Fisheries and Oceans Canada
Ecosystems and Oceans Science

Pêches et Océans Canada
Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2016/035

Pacific Region

Methodology of the Pacific Marine Ecological Classification System and its Application to the Northern and Southern Shelf Bioregions

Emily Rubidge[1], Katie S. P. Gale[1], Janelle M. R. Curtis[2], Erin McClelland[3], Laura Feyrer[4], Karin Bodtker[5], Carrie Robb[5]

[1]Institute of Ocean Sciences
Fisheries & Oceans Canada
P.O. Box 6000
Sidney, BC V8L 4B2

[2]Pacific Biological Station
Fisheries & Oceans Canada
3190 Hammond Bay Rd
Nanaimo, BC V9T 1K6

[3]EKM Scientific Consulting

[4]BC Ministry of Environment
P.O. Box 9335 STN PROV GOVT
Victoria, BC V8W 9M1

[5]Living Oceans Society
204-343 Railway St.
Vancouver, BC V6A 1A4

May 2016

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

© Her Majesty the Queen in Right of Canada, 2016
ISSN 1919-5044

Correct citation for this publication:

Rubidge, E., Gale, K.S.P., Curtis, J.M.R., McClelland, E., Feyrer, L., Bodtker, K., and Robb, C. 2016. Methodology of the Pacific Marine Ecological Classification System and its Application to the Northern and Southern Shelf Bioregions. DFO Can. Sci. Advis. Sec. Res. Doc. 2016/035. xi + 124 p.

TABLE OF CONTENTS

ABSTRACT x

RÉSUMÉ xi

1 INTRODUCTION 1

- 1.1 HIERARCHICAL MARINE ECOLOGICAL CLASSIFICATION IN THE PACIFIC REGION 1
- 1.2 MANAGEMENT NEEDS FOR A PACIFIC MARINE ECOLOGICAL CLASSIFICATION SYSTEM (PMECS) 1
  - 1.2.1 Marine Protected Area planning and ecological representativity 1
- 1.3 EXISTING MARINE CLASSIFICATIONS IN BRITISH COLUMBIA 2
  - 1.3.1 The British Columbia Marine Ecological Classification System (BCMEC) 2
  - 1.3.2 ShoreZone database 4
- 1.4 PMECS DEVELOPMENT 4
- 1.5 CONTEXT AND PROJECT OBJECTIVES 8
- 1.6 APPROACH 10
  - 1.6.1 Modifications to conceptual framework to move towards PMECS implementation 10

2 DATA SOURCES 13

- 2.1 DEVELOPMENT OF THE PMECS GEOSPATIAL DATABASE 13
- 2.2 BIOTIC DATA 13
  - 2.2.1 OBIS 14
  - 2.2.2 Royal BC Museum (RBCM) 14
  - 2.2.3 Shellfish Data Holdings: SF Log and SF Bio 14
  - 2.2.4 Groundfish Biological Survey Database (GF Bio) 14
  - 2.2.5 Biotic data preparation 15
- 2.3 ABIOTIC DATA 15
  - 2.3.1 Depth and derivatives 16
  - 2.3.2 Chlorophyll A 16
  - 2.3.3 Sea surface temperature 16
  - 2.3.4 Bottom Salinity, temperature, currents 16
  - 2.3.5 Habitat-Template model 17
  - 2.3.6 Grain size model 17
  - 2.3.7 Nutrients 17

3 LEVEL 4A AND B: BIOPHYSICAL UNITS 19

- 3.1 APPROACH 19
- 3.2 METHODOLOGY AND APPLICATION 20
  - 3.2.1 Selection of data 20
  - 3.2.2 Cluster Analysis Methodology 21
  - 3.2.3 Indicator species analysis 25
  - 3.2.4 Environmental correlates of biological clusters 25

3.3 CLUSTER ANALYSIS RESULTS 30
3.3.1 Description of clusters 30
3.3.2 Environmental differences among biophysical units 38
3.4 LEVEL 4A AND B: BIOPHYSICAL UNITS DISCUSSION 44
3.4.1 Model performance, uncertainty and field validation 45
3.4.2 Limitations of the approach 46
3.4.3 Conclusions 48
4 LEVEL 5: GEOMORPHIC UNITS 48
4.1 GEOMORPHIC UNIT APPROACH 48
4.2 METHODOLOGY AND APPLICATION 49
4.2.1 Data 49
4.2.2 Summary of methods 49
4.2.3 Bathymetry correction 50
4.2.4 Analysis area delineation 50
4.2.5 Bathymetric Position Index parameters 52
4.2.6 Classification rules 53
4.3 GEOMORPHIC UNITS RESULTS 54
4.3.1 Bathymetric Position Index parameters 54
4.3.2 Classification rules 55
4.3.3 Post-processing of classified Geomorphic units 57
4.4 LEVEL 5: GEOMORPHIC UNITS: LESSONS, LIMITATIONS, AND FUTURE CONSIDERATIONS DISCUSSION 61
5 LEVEL 6: BIOTOPES 62
6 LEVEL 7: BIOLOGICAL FACIES 64
7 GENERAL DISCUSSION 65
7.1 CLIMATE CHANGE AND ECOLOGICAL CLASSIFICATION 66
7.2 HIERARCHICAL CLASSIFICATION SYSTEMS AND ISSUES OF SCALE 66
7.3 USE OF SPATIAL AUTOCORRELATION PATTERNS 67
7.4 INFORMATION GAPS 68
7.5 SUMMARY AND FURTHER WORK 68
8 ACKNOWLEDGEMENTS 71
9 DEFINITIONS 71
9.1 GLOSSARY 71
9.2 ACRONYMS 71
10 LITERATURE CITED 72
APPENDIX A. BIOLOGICAL RELEVANCE OF BCMEC ECOSECTIONS 81
METHODS 81
RESULTS 81
APPENDIX B. DETAILS ON PRIMARY DATA COLLECTION AND PREPARATION 86
SHELLFISH DATA HOLDINGS (SF LOG AND SF BIO) 86

Crab by Trap 88
Tanner Crab 88
GROUNDFISH DATABASES (GFBIO) 89
Hake Stock Delineation 90
Hecate Strait Multispecies Trawl Survey 90
Hecate Strait Pacific Cod Monitoring Trawl 91
Inshore Rockfish Longline Survey 91
International Pacific Halibut Commission (IPHC) Longline Survey 92
Joint Canada/US Hake Acoustic Survey 93
Pacific Halibut Management Association (PHMA) Longline Survey 93
Sablefish Research and Assessment Survey 94
Synoptic Surveys (Hecate Strait, Queen Charlotte Sound, West Coast Vancouver Island, West Coast Queen Charlotte Island) 95
West Coast Vancouver Island Thornyhead Trawl Survey 97
Yelloweye Rockfish Charter Longline Survey 97
APPENDIX C. CORRELATION ANALYSIS 98
APPENDIX D. SPECIES INFORMATION 100
APPENDIX E. TRANSITION ZONE DIVERSITY ANALYSIS 117
METHODS 117
LEVEL 4A BIOPHYSICAL UNITS 117
LEVEL 4B BIOPHYISCAL UNITS 117
APPENDIX F. GEOMORPHIC UNITS 121
GEOMORPHIC UNIT ATTRIBUTES 121
LITERATURE CITED - APPENDICES 123

LIST OF TABLES

Table 1. BCMEC divisions in Pacific Region, from top of hierarchy (Ecozone) to bottom (Ecounit) Source: Harper et al. (1993), Zacharias et al. (1998), AXYS Environmental Consulting Ltd. (2000, 2001). 3

Table 2. Prototype design of the hierarchical Pacific Marine Ecological Classification System (PMECS) developed at the Pacific regional peer review Meeting 12–14 Feb. 2013. Source: DFO (2013). 6

Table 3. Summary of recommendations from CSAS Science Advisory Report (DFO 2013) that were considered in this working paper while developing the methodologies to apply PMECS to NSB. 7

Table 4. Levels, definitions, data needs, outputs and classification methods of the hierarchical Pacific Marine Ecological Classification System framework. 11

Table 5. List of environmental data layers compiled and/or generated for use in environmental analysis 18

Table 6. Indicator species for Level 4a (large-scale Biophysical units) clusters produced by cluster analysis of species assemblages of demersal fish and invertebrates listed in order of Indval metric. 33

Table 7. Indicator species for Level 4b clusters produced by cluster analysis of species assemblages of demersal fish and invertebrates listed in order of Indval metric. 37

Table 8. Features measured in each analysis area, range of sizes, and final scale factors chosen. 54

Table 9 Classification table for continental slope analysis area. 55

Table 10 Classification table for continental shelf (without fjords) analysis area. 56

Table 11. Classification table for fjords analysis area. 57

Table 12. Description of data sets used in Gregr et al. (2013) to classify nearshore benthic ecosystems (summarized from Gregr et al. (2013). 62

Table 13. Summary of recommendations from CSAS Science Advice Report (DFO 2013/065) and how these were addressed in our working paper. 69

Table A 1. PERMANOVA analysis of variance table for Ecosections analysis, including 5 Ecosections (Continental Slope, Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf). 83

Table A 2. PERMDISP analysis of variance table for Ecosections analysis, including 5 Ecosections (Continental Slope, Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf). 83

Table A 3. PERMANOVA analysis of variance table for Ecosections analysis, including 4 Ecosections (Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf). 83

Table A 4. PERMANOVA analysis of variance table for Level 4b biophysical unit analysis. 83

Table A 5. PERMDISP analysis of variance table for Level 4b biophysical unit analysis. 83

Table B 1. Description, years, and number of records and taxa from databases within the DFO Shellfish Log (SF Log; commercial fisheries) data holdings. 86

Table B 2. Description, years, and number of georeferenced records and taxa from databases within the DFO Bio (SF Bio; research) data holdings. 87

Table D 1. Species included in cluster analysis for Level 4a and 4b biophysical unit identification. 100

Table D 2. List of species considered for use in cluster analysis, but excluded due to low frequency across study sites 105

Table D 3. List of species excluded from analysis because of a pelagic, freshwater, or unknown habitat type 114

Table E 1. ES_n (expected number of species) for biophysical unit sites (random forest probability of assignment ≥ 0.70), and transition sites (probability of assignment < 0.70), where n is the number of sites in either biophysical unit or transition sections. 118

Table F 1. Classification scheme for geological features and zones after Manson (2009). 121

Table F 2. Classification scheme for geological features and zones after Ferdaña (2006). 122

LIST OF FIGURES

Figure 1. BCMEC divisions. A) Ecoprovinces, B) Ecoregions, C) Ecosections, and D) Ecounits. 3

Figure 2. Marine Bioregions of Canada's Pacific Region. 8

Figure 3. Map of study area and sample sites. The study area includes 6875 sites (4 km grid cells) within the Northern Shelf and Southern Shelf Bioregions (total area is 110,000 km^2). 23

Figure 4. Number of sites (out of a possible 3615) in each of the top 8 clusters for dendrograms cut at increasing β_{sim} values. 24

Figure 5. Variance (standard deviation) of the number of sites in the top 8 clusters compared to the number of sites retained in those clusters for increasing β_{sim} cut-off values. β_{sim} values of 0.55 (square) and 0.65 (circle) show variance and number of sites for the biophysical unit analyses, respectively. 25

Figure 6. Sampling effort for each sampling site (4 km grid cell) for data included in the cluster analysis and random forest analysis. 27

Figure 7. Number of species recorded at each sampling site (4 km grid cell) for the dataset used in the cluster analyses. 28

Figure 8. Model fit (AUC) for random forest models predicting membership to clusters resulting from dendrograms cut at different β_{sim} values. 30

Figure 9. Dendrogram (top) cut at β_{sim} height of 0.65 and resulting map (bottom) of sites belonging to each cluster. 32

Figure 10. Dendrogram (top) cut at β_{sim} height of 0.55 and resulting map (bottom) of sites belonging to each cluster. 36

Figure 11. Variable importance (mean ± SD decrease in accuracy) for random forest model (ten runs of random 70% of data, 10,000 trees each run) for whole model and for each large-scale biophysical unit individually. 38

Figure 12. Boxplots of most important environmental variables for differentiating the three large-scale biophysical units, as determined by a random forest analysis. 39

Figure 13. Random forest model-predicted output for site classification into major clusters (Level 4a, large-scale Biophysical units). 40

Figure 14. Level 4a (large-scale Biophysical units) classification showing sites with low (< 0.70) probability of assignment from the random forest analysis. 41

Figure 15. Variable importance (mean ± SD decrease in accuracy) for random forest model (ten runs of random 70% of data, 10,000 trees each run) for whole model and for each biophysical unit individually. 42

Figure 16. Boxplots of most important environmental variables for differentiating the five biophysical units, as determined by a random forest analysis. 42

Figure 17. Random forest model-predicted output for site classification into major clusters (Level 4b, biophysical units). 43

Figure 18. Level 4b (biophysical units) classification showing sites with low (< 0.70) probability of assignment from the random forest analysis. 44

Figure 19. Map of study area for geomorphic unit analysis. 51

Figure 20. Top: with a small scale factor, subtle topographic features can be delineated. 53

Figure 21. Continental slope base, newly defined continental shelf break with 200 and 500 m depth contours. 58

Figure 22. Final geomorphic units for three analysis areas. 60

Figure 23. Example of Bottom Patch Classification: a and b show the bottom type while c and d show the confidence level of those classifications. 63

Figure 24. Glass sponge reefs as examples of biological facies. 65

Figure A 1. Subset of British Columbia Marine Ecological Classification (BCMEC) Ecosections analyzed in this paper. 84

Figure A 2. Non-metric multidimensional scaling plot of A) BCMEC Ecosections and B) Level 4b biophysical units, with 95% confidence ellipses. 85

Figure B 1. Hake Stock Delineation Survey locations. 90

Figure B 2. Hecate Strait Multispecies Trawl locations. 90

Figure B 3. Hecate Strait Pacific Cod Monitoring Trawl locations. 91

Figure B 4. Inshore Rockfish Longline Survey locations. 92

Figure B 5. International Pacific Halibut Commission (IPHC) Longline Survey locations. 93

Figure B 6. Joint Canada/US Hake Acoustic Survey locations. 93

Figure B 7. Pacific Halibut Management Association (PHMA) Longline Survey locations. 94

Figure B 8. Sablefish Research and Assessment Survey locations. 95

Figure B 9. Synoptic Survey locations (Clockwise from upper left: Hecate Strait, Queen Charlotte Sound, West Coast Vancouver Island, West Coast Queen Charlotte Island) 96

Figure B 10. West Coast Vancouver Island Thornyhead Trawl Survey locations. 97

Figure B 11. Yelloweye Rockfish Charter Longline Survey locations. 97

Figure C 1. Correlation plot of 59 environmental variables at 4-km resolution. 98

Figure C 2. Correlation plot of final 14 environmental variables that were used in the random forest analyses. 99

Figure E 1. Species accumulation curves for the level 4a biophysical unit analysis. 119

Figure E 2. Species accumulation curves for the level 4b biophysical unit analysis. 120

ABSTRACT

Marine ecological classification systems are needed to place species, habitats and ecosystems at varying spatial scales into ecological and management contexts. In Canada, the need to develop a hierarchical ecological classification system has been recognized at national and regional scales. The Pacific Marine Ecological Classification System (PMECS) conceptual framework to disaggregate the four major bioregions in Canada's Pacific Region into smaller spatial units was developed over the past several years (2009–2013). In this paper, we advance PMECS towards implementation by developing analytical methods to populate the benthic hierarchical layers of the conceptual PMECS framework proposed by Robinson et al. (2015). Six specific objectives were met in this working paper to move PMECS from a conceptual framework to implementation. First, we compiled existing spatially referenced biotic and abiotic data for use in marine spatial planning in Canada's Pacific Region into a central geodatabase. Second, we selected appropriate datasets from the geodatabase and developed a biologically driven community approach using both abiotic and biotic data to generate a broad-scale classification for the Northern Shelf and Southern Shelf Bioregions (NSB, SSB). Third, we applied this community approach to the entire NSB and SSB to populate Levels 4a and b of PMECS framework (Biophysical units). Fourth, we used an indicator species analysis to identify species that were most commonly associated with each Biophysical unit. Fifth, we developed and applied classification methods to the NSB/SSB to delineate Geomorphic units (Level 5) using a 75 m bathymetry raster and the benthic positioning index (BPI) in the Benthic Terrain Modeller toolbox. Sixth, we proposed quantitative methods suitable for classifying areas at finer scales (lower levels) of PMECS. We recommend investigating the bottom patch model (Gregr et al. 2013) to classify and map substrate types and evaluating its application to populate Level 6 of PMECS (Biotopes) as more data become available. Known distributions of Sponge reefs, Kelp forests, and other Biological Facies (Level 7) can be incorporated into PMECS with the understanding that these units may be incompletely mapped in Pacific Region. Classification results at Levels 4 and 5 are recommended for use in marine protected area (MPA) network design, and classifications at Levels 6-7 are proposed for integration into MPA network design as these analyses are completed within Bioregions.

Méthodes d'application du système de classification de l'écologie marine du Pacifique aux biorégions des plateaux nord et sud

RÉSUMÉ

Les systèmes de classification de l'écologie marine sont nécessaires pour appliquer les espèces, les habitats et les écosystèmes à diverses échelles spatiales dans des contextes de gestion et d'écologie. Au Canada, la nécessité d'élaborer un système de classification hiérarchique de l'écologie a été reconnue à l'échelle régionale et nationale. Au cours des dernières années (2009-2013), on a élaboré le cadre conceptuel du système de classification de l'écologie marine du Pacifique (SCEMP) afin de découper les quatre principales biorégions de la Région du Pacifique au Canada en plus petites unités spatiales. Dans le présent document, nous progressons vers la mise en œuvre du SCEMP en élaborant des méthodes analytiques visant à remplir les couches hiérarchiques benthiques du cadre conceptuel du SCEMP proposé par Robinson *et al.* (2015). Six objectifs précis ont été atteints dans le présent document de travail pour faire du cadre conceptuel du SCEMP un cadre pratique. Premièrement, nous avons compilé les données biotiques et abiotiques à référence spatiale existantes aux fins de planification spatiale marine dans la Région du Pacifique du Canada dans une base de données géoréférencées centrale. Deuxièmement, nous avons sélectionné les ensembles de données appropriés dans cette base de données, puis élaboré une approche biologique axée sur la communauté au moyen des données biotiques et abiotiques afin d'obtenir une classification à grande échelle pour les biorégions des plateaux nord et sud. Troisièmement, nous avons appliqué cette approche communautaire à l'ensemble des biorégions des plateaux nord et sud pour remplir les niveaux 4a et 4b du cadre du SCEMP (unités biophysiques). Quatrièmement, nous avons mené une analyse des espèces indicatrices pour identifier les espèces les plus souvent associées à chaque unité biophysique. Cinquièmement, nous avons élaboré et appliqué des méthodes de classification dans les biorégions des plateaux nord et sud pour délimiter les unités géomorphologiques (niveau 5) au moyen d'une trame bathymétrique de 75 m et de l'indice de position benthique dans l'outil de modélisation du terrain benthique. Sixièmement, nous avons proposé des méthodes quantitatives appropriées pour classifier les zones à de plus petites échelles (niveaux inférieurs) du SCEMP. Nous recommandons d'étudier le modèle de parcelles de hauts-fonds (Gregr *et al.* 2013) pour classifier et cartographier les types de substrats et évaluer son application de remplissage du niveau 6 (biotopes) du SCEMP à mesure que d'autres données deviennent disponibles. Les répartitions connues de récifs d'éponges, de forêts de varech et d'autres faciès biologiques (niveau 7) peuvent être intégrées dans le SCEMP à condition de comprendre que la cartographie de ces unités peut être incomplète dans la Région du Pacifique. Il est recommandé d'utiliser les résultats de la classification aux niveaux 4 et 5 lors de la conception du réseau de zones de protection marine (ZPM), et les classifications aux niveaux 6 et 7 peuvent être intégrées dans la conception des ZPM, car ces analyses sont effectuées dans les biorégions.

1 INTRODUCTION

1.1 HIERARCHICAL MARINE ECOLOGICAL CLASSIFICATION IN THE PACIFIC REGION

The purpose of ecological classifications is to map an area into relatively homogenous units for ecological research, monitoring, and management. Marine ecological classification systems are needed to place species, habitats and ecosystems at varying spatial scales into ecological and management contexts and are used to underpin marine spatial planning. In Canada, the need to develop a hierarchical ecological classification system has been recognized at national and regional levels. Although 12 major biogeographic units (also referred to as Bioregions; DFO 2009) have been identified in Canada's three oceans, these Bioregions need to be further subdivided in an ecologically meaningful way to provide the information necessary to inform bioregional policy development and management decisions (DFO 2009).

1.2 MANAGEMENT NEEDS FOR A PACIFIC MARINE ECOLOGICAL CLASSIFICATION SYSTEM (PMECS)

At full implementation, the Pacific Marine Ecological Classification System (PMECS) will have broad application to varying marine spatial planning management needs, from conservation planning to fisheries management and environmental impact assessment. Overall, a standardized, ecologically based classification system will reduce the uncertainty about where species, habitats and ecosystems occur, thereby leading to more informed decision making across management sectors. The initial focus and purpose of this paper is to develop the benthic component of PMECS to underpin marine protected area (MPA) network design. However, once the classification system is complete for both the pelagic and the benthic components, it will have broad utility for other marine spatial planning initiatives, particularly when the finer levels of PMECS are populated for the entire Pacific region.

1.2.1 Marine Protected Area planning and ecological representativity

A key commitment under Canada's *Ocean's Act* and a part of Canada's Oceans Strategy is the development of a national MPA network. Design criteria for establishing an MPA network include ecological representativity and the protection of biodiversity and special natural features (Canada – British Columbia Marine Protected Area Network Strategy 2014). Representativity is defined as "relatively intact, naturally functioning examples of the full range of ecosystems and habitat diversity found within a given planning area (Canada – British Columbia Marine Protected Area Network Strategy 2014).To ensure all types of ecosystems are effectively represented, there is a need to use the best available data to characterize the structure and distribution of Canada's marine biota and habitats across multiple spatial scales.

The MPA network design criterion of representativity builds ecological resilience into the network to buffer the impacts of human activities by incorporating a proportion of each type of ecosystem and habitat (Sheppard 2013). Achieving representativity in an MPA network necessitates information on the distribution of ecosystems, including the distribution of species and habitats across multiple spatial scales (Roff et al. 2003, Last et al. 2010, Roff and Zacharias 2011, Harris 2012a, Sheppard 2013, Robinson et al. 2015). An ecological classification system that partitions areas into relatively homogeneous spatial units based on a selected set of environmental and/or biological variables can be used to delineate ecological units as a basis for implementing the representativity criterion (Roff and Zacharias 2011).

There are a variety of approaches used to classify ecosystems from global to local scales (e.g., Madden et al. 2005, Valentine et al. 2005, Spalding et al. 2007, Greene et al. 2008, UNESCO 2009). All classification approaches result in a simplified categorization of an area whether they are defined by differences in ecological processes (e.g., McMahon et al. 2004), environmental conditions (e.g., Bailey et al. 1985, Roff et al. 2003), or biological communities (e.g., Olson et al. 2001, Allen and Pondella 2006). Ideally, marine ecological classifications should reflect the relationships between physical features and the distribution and abundance of species (Gregr et al. 2012), however due to limitations in the availability of biological data, most marine ecological classifications are built on the physical geography or the abiotic conditions of the planning area. These physiographic classifications assume that the physical variables are reliable surrogates of biological patterns as well (e.g., Roff et al. 2003).

While abiotic surrogates can reflect biological patterns (e.g., Roff and Taylor 2000, Bredin et al. 2001), physiographic classifications do not perform as well as biologically informed classifications at fulfilling the representativity criterion in conservation planning (Rodrigues and Brooks 2007, Ban 2009, Sutcliffe et al. 2015). Biological representativity has shown to be low in reserves designed on abiotic domains alone (Lombard et al. 2003, Sutcliffe et al. 2015). Sutcliffe et al. (2015) found that abiotic classifications may be used for an initial reserve design when biological information is insufficient, but classifications that are biologically informed, either through weighting the biological importance of the abiotic variables (e.g., Pitcher et al. 2012) or by explicitly incorporating biological data, produce more representative reserves. These results highlight the importance of integrating available biological data into the ecological classification to achieve ecological representativity in MPA networks.

1.3 EXISTING MARINE CLASSIFICATIONS IN BRITISH COLUMBIA

The purpose of this section is not to review marine ecological classification systems (see Madden and Grossman 2004, Harris 2012a), but to provide an overview of two classification systems that currently exist in the Pacific Region as context for this working paper.

1.3.1 The British Columbia Marine Ecological Classification System (BCMEC)

BCMEC was developed for the Pacific Coast of Canada in the 1990s with some additions and revisions over the decade (Harper et al. 1993, Zacharias et al. 1998, AXYS Environmental Consulting Ltd. 2000, 2001). The classification is hierarchical and consists of five nested divisions based on physical properties. Table 1 and Figure 1 below briefly describe the data used to define each level in the BCMEC hierarchy.

Table 1. BCMEC divisions in Pacific Region, from top of hierarchy (Ecozone) to bottom (Ecounit) Source: Harper et al. (1993), Zacharias et al. (1998), AXYS Environmental Consulting Ltd. (2000, 2001).

Level name	Criteria	BCMEC classes
Ecozone	Ice regimes and ocean basins	Pacific
Ecoprovince	Ocean regimes and continental margins	North Pacific, Pacific Shelf and Mountains, Georgia-Puget Basin
Ecoregion	Marginal seas	Subarctic Pacific, Transitional Pacific, Outer Pacific Marine Shelf, Inner Pacific Marine Shelf, Inner Pacific Marine Shelf, Georgia Basin
Ecosection	Mixing and stratification	*Examples*: Continental Slope, Vancouver Island Shelf, Queen Charlotte Sound, Dixon Entrance, Hecate Strait, North Coast Fjords, etc.
Ecounit	Current, depth, bottom substrate, bottom relief, wave exposure	Unique combinations of each criterion

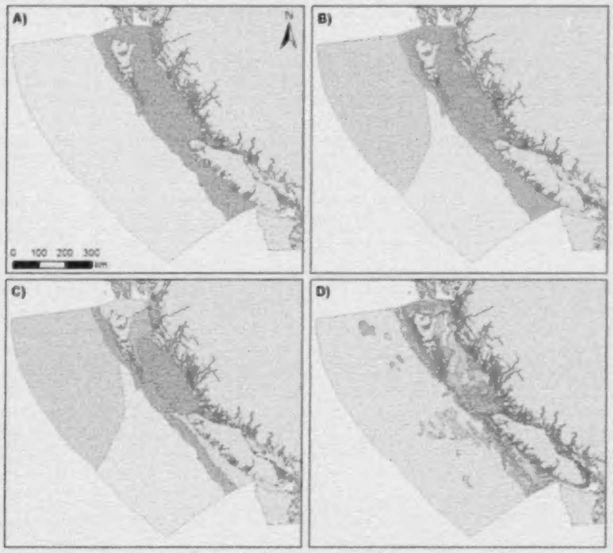


Figure 1. BCMEC divisions. A) Ecoprovinces, B) Ecoregions, C) Ecosections, and D) Ecounits.

The fifth division, the "Ecounit" was developed for benthic and pelagic systems because the "Ecosection" level was deemed too large spatially for use in coastal management and MPA planning (Zacharias et al. 1998). In addition, the Ecosections were generated using a combination of expert opinion and limited data, but there was no method of verifying the accuracy of their boundaries so Zacharias et al. (1998) added the Ecounit to the hierarchy of BCMEC with the objective of further delineating the Pacific Region, and evaluating the boundaries of the Ecosections.

Although the Ecounit was developed for the purposes of coastal management and marine protected area planning, it has not been readily used in these processes. In general, BCMEC, despite the considerable amount of work that went into it, has faced two main criticisms:

1. a lack of biological criteria in operationally defining the units, and
2. the accuracy of Ecounit boundaries.

The Ecounit, despite some revisions (see AXYS Environmental Consulting Ltd. 2000, 2001), is criticized based on its resolution (scale and accuracy of input substrate layer), methodology (not repeatable) and biological relevance (Levings and Jamieson 1999, Johannessen et al. 2004). Neither the BCMEC Ecosections nor the Ecounits had been validated using biological information so their suitability for analyses of ecological representativity in MPA network planning is uncertain (although see Appendix A for biological validation of Ecosections).

1.3.2 ShoreZone database

"ShoreZone" (Howes et al. 1994) is an intertidal habitat classification derived from low tide aerial surveys on the coast of British Columbia. These surveys recorded both geological features and biological zonation and were used to classify shore units based on coastal classes (e.g., gravel flat, sand beach, rock platform) and "biobands" (e.g., barnacles, mussels, different types of algae). Despite it being somewhat outdated (data collected more than 25 years ago in certain areas) ShoreZone continues to be the best available intertidal habitat classification in the Pacific Region.

1.4 PMECS DEVELOPMENT

A Pacific Region science advisory process in February 2013 reviewed a conceptual framework for the Pacific Marine Ecological Classification System (PMECS) for use as a tool to provide spatially explicit information on marine biotic and abiotic diversity (DFO 2013, Robinson et al. 2015). The advisory process concluded that application of the conceptual framework (shown in Table 2) could be used to help describe biodiversity patterns and aid in the ecosystem based management of marine resources on the BC coast (DFO 2013). To move towards implementation of PMECS in the Pacific Region, the conceptual framework needed to be further developed by identifying available data, proposing analytical procedures and applying the classification system in the region. Robinson et al. (2015), upon review of existing provincial, national and international ecological classifications, provided a hierarchical prototype of PMECS that was based primarily on the Australian seabed bioregionalization framework (Last et al. 2010), with modifications specific to Canada's Pacific Region (DFO 2013). This classification framework differs from other systems for classifying marine biota by explicitly recognizing the overarching influence of large-scale biodiversity patterns at the realm, provincial and bathymetric (depth-related) levels (Lyne et al. 2006, Last et al. 2010), ensuring that the large-scale patterns of biodiversity in these higher levels are captured in lower levels of the hierarchy. The Science Advisory Report resulting from the review of the conceptual PMECS framework resulted in fourteen recommendations summarized in Table 3 (DFO 2013). Robinson et al. (2015) also identify four key considerations when modifying the Last et al. (2010) classification framework for application in Canada's Pacific Region: First, they suggest that the classification system needs to be developed with ecosystems in mind; second, it needs to be based on a classification system that is currently used for marine resource management (i.e., Last et al. 2010 in Australia); third, the classification system should be hierarchical so that it provides information at nested spatial extents relevant to managers; and finally, the classification system should include a strong biological component.

In this paper, we draw on the recommendations outlined in Robinson et al. (2015) and the science advisory report (DFO 2013,Table 3) and use them to further develop and modify the conceptual PMECS framework and develop some of the methods needed to apply it to Canada's Pacific Region (Figure 2). While pelagic data were compiled for future classification of pelagic ecosystems, this working paper focuses on methodology, analyses and classifications of **benthic** ecosystems at multiple spatial scales. Specifically, this working paper aims to support the development of science advice in a zonal peer review process on hierarchical marine ecological classification systems for Pacific and Maritimes Regions. The peer review process objectives are threefold:

1. Review the methods and results of the application of the conceptual classification framework in each Region, given data availability and their intended purposes;
2. Discuss uncertainties and consequences associated with data availability and classification decisions (e.g., number, boundary, type, etc.) made in the Fisheries and Oceans Canada (DFO) Pacific and DFO Maritimes application of the conceptual framework, and provide guidance for future application; and,
3. Provide guidance on appropriate types of analyses to classify areas at spatial scales not completed as part of these applications.

Table 2. Prototype design of the hierarchical Pacific Marine Ecological Classification System (PMECS) developed at the Pacific regional peer review Meeting 12–14 Feb. 2013. Source: DFO (2013).

Level	Unit	Spatial extent	Spatial resolution	Benthic description	Pelagic description
0	Realm	10,000's km	1,000 km^2	Broad-scale geographic units such as the north Pacific Ocean.	
1	Province	1,000's km	~ 100 km^2	Broad-scale geological units such as continental blocks, basins and abyssal plains.	Zoogeographic provinces (e.g., Oregonian, Aleutian).
2	Bioregions	1,000's km	~10-100 km^2	Distinctive, recurring and small-scale physical oceanographic processes (e.g., separation between California Current and Alaska Current regions). Four major bioregions in Pacific marine waters. Research and analysis is required to understand how marine species diversity differs among these Bioregions.	
3	Ecosections	100's-1,000's km	~10-100 km^2	Ecosections are primarily related to abiotic pelagic oceanographic processes; relation to benthic ecosystems requires further research.	Distinct, recurring, and large-scale physical oceanographic processes and topographic or bathymetric features. For example, the Vancouver Island Coastal Current and the Juan de Fuca Eddy.
4	Bathomes	100's-1,000's km	~10 km^2	Nearshore and littoral zone, continental shelf, continental slope, abyssal plain.	Neritic zone, epipelagic zone, mesopelagic zone, bathypelagic zone, abyssopelagic zone.
5	Geozones	100's km	1-10 km^2	Mappable areas with similar seabed geomorphology and usually with distinct biota (e.g., seamounts, canyons, rocky banks, inlets).	Mappable structures based on oceanographic processes assumed to be surrogates for distinctive biological assemblages (tidal mixing areas, fronts, upwellings).
6	Primary biotopes	10's-100's km	< 1 km^2	Nested within Geomorphic Units are soft, hard or mixed substrate-based units, together with their associated substrate-based units and their associated biological communities.	Combinations of physical and chemical water property data (sea surface temperature and salinity, dissolved oxygen, stratification) and associated biological communities.
7	Secondary biotopes	100's-1000's m	100's m^2	Smaller-scale abiotic and biotic sub structural units characterized by specific types of substrate (e.g., seapen beds, sponge reefs).	Detailed combinations of physical data to describe water masses (e.g., chlorophyll maxima, pycnocline).
8	Biological facies	100's m	< 10 m^2	Fundamental unit for management of biodiversity. Mappable units that act as surrogates for all levels below (e.g., species of seagrass, group of hard corals or sponges).	Pelagic (mobile) taxa less informative descriptors of facies than sessile plants and animals.
9	Micro-communities	10's m	< 1 m^2	Assemblages of species that depend on member species of the Biological Facies, e.g., holdfast communities in giant kelp.	
10	Species	Discussion of the hierarchy did not proceed to these lower levels as their descriptions are inconsistent with the hierarchical nature of the benthic and pelagic descriptions of higher levels			
11	Populations				
12	Genes				

Table 3. Summary of recommendations from CSAS Science Advisory Report (DFO 2013) that were considered in this working paper while developing the methodologies to apply PMECS to NSB.

SAR Recommendation	Addressed in this working paper?
1. Classification system must be hierarchical with multiple levels spanning bioregions to micro-communities (i.e., multiple and nested spatial scales).	Yes. Spatial scales are nested in the hierarchy but not all ecological units are necessarily nested due to natural variation (i.e. nature is not perfectly nested).
2. Benthic and pelagic classification systems are needed.	No, initial quantitative analyses focus on **benthic** classification
3. PMECS needs to be created and evaluated with respect to management objectives at varying spatial scales.	Yes
4. Application of a suite of tools to analyze and summarize biotic and abiotic data	Yes
5. Identification of important data sources and gaps.	Yes
6. Uncertainty must be captured and documented (including the drivers of uncertainty) at each level in the framework.	Yes
7. Guidance is needed on how to address some known uncertainties/error in fundamental data sources (i.e., acoustic bathymetry and bottom type, water properties, abiotic and biotic surveys, and satellite, photographic and video imagery). This guidance will document the procedure for uncertainty documentation.	Yes
8. Guidance needs to be developed for the process for incorporating new data as they become available.	Once framework has been accepted, guidance for new data can be developed.
9. The performance of the PMECS prototype will be tested with existing data and evaluated with the appropriate metrics.	Yes
10. PMECS should make use of DFO's ongoing climate change research projections and trends.	Yes
11. A parallel pilot program should be developed to evaluate the performance of different methods using the same datasets to provide guidance on model choices to PMECS users.	No
12. Collaboration is a critical component of PMECS development. PMECS development is analyses heavy and requires multiple data sources, technical expertise and data sharing. It depends on numerous partners.	Yes
13. There is a recognized gap in the coordination of GIS capacity for geospatial planning and management internally within DFO and externally with other agencies/partners. Standards for data collection, storage and sharing will be needed as implementation of PMECS proceeds.	Yes
14. The development of PMECS, following the guidance and advice given during the PMECS Regional Peer Review, will result in progress towards fulfilling national commitments to ensure that adequate MPA network design is achieved in the Pacific Region.	Yes

1.5 CONTEXT AND PROJECT OBJECTIVES

In this study, we advance PMECS towards implementation by developing analytical methods, and integrating classifications that are currently being developed in the region (e.g., Gregr et al. 2013), to populate the hierarchical layers of the conceptual PMECS framework proposed by Robinson et al. (2015). Given that BCMEC, a physiographic classification system, already exists but is not readily used in Pacific marine spatial planning initiatives, and the evidence suggesting that incorporating biological data into ecological classifications results in more representative reserves (Lombard et al. 2003, Sutcliffe et al. 2015), we focus on a biological approach to populate the higher levels of PMECS. This also ensures larger-scale biodiversity patterns are captured at lower levels in the hierarchy (Table 4) as discussed by Last et al. (2010). Once complete, PMECS will have broad application to varying management needs, from fisheries management and environmental impact assessment to aquaculture siting. However, the immediate need is to develop a hierarchical classification system to underpin MPA network design in Pacific Region, with an initial focus on the benthic environment of the Northern Shelf Bioregion (NSB; Figure 2). Here, we classify the NSB and the Southern Shelf Bioregion (SSB) given the scale of the analysis and recognizing that a classification will be needed in both regions.

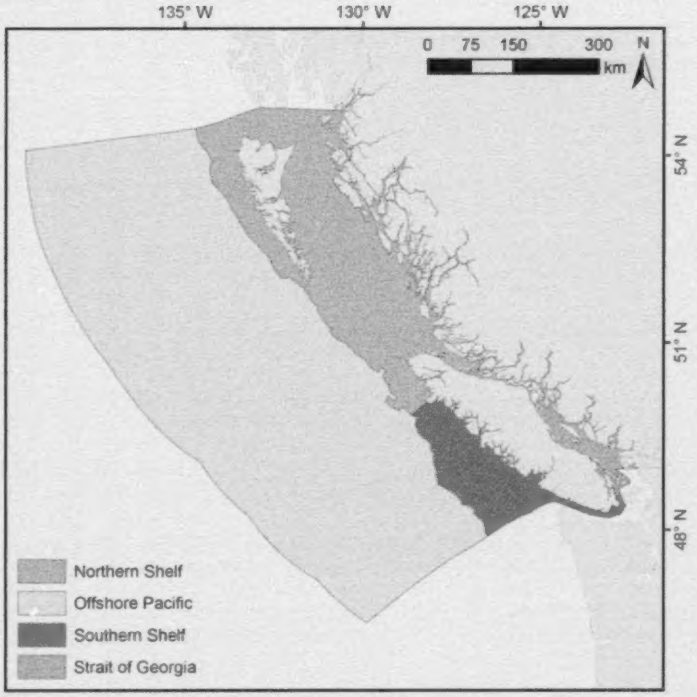


Figure 2. Marine Bioregions of Canada's Pacific Region.

The Canada – British Columbia Marine Protected Area Network Strategy (2014) defines their goal of representativity as "protecting relatively intact, naturally functioning examples of the full range of ecosystems and habitat diversity found within a given planning area". Implicit in this goal are representative areas of both the biological and physical diversity that supports marine ecosystems and habitats. There have been efforts to produce benthic habitat representativity maps in the Pacific Region, but none have explicitly included biotic data into the classification of representative ecological units, mainly due to data constraints. One example of this is an effort by Parks Canada to use physical features to classify the habitats in the Strait of Georgia (Robinson and Royle 2008) with the assumption that the physical habitat diversity acts as a surrogate for biodiversity. Although this assumption has support in the literature, particularly for substrate type (e.g., Stevens and Connolly 2004, Post et al. 2006), the authors acknowledged the need to test this assumption. The relationships between physical variables and the occurrence and abundance of benthic species has been the topic of considerable research (see Harris 2012b for review) and most marine species are limited by some combination of depth, temperature, salinity and substrate type (Roff and Zacharias 2011). However, the complexity of the relationships are less well understood and further complicated by biotic interactions (see Harris 2012b for review).

In 2012, a national science review process in response to the DFO Program Policy Sector requesting science advice on how to incorporate representativity into MPA network planning resulted in several recommendations on how to ensure the consistency of scale at which representativity must be considered (DFO 2012b). The number of hierarchical levels of classification needed to inform conservation and management decisions is dependent upon the spatial extent and resolution required to achieve specific objectives. Scale of analysis is also influenced by the physical and biological heterogeneity of the marine environment and limited by data availability. The national science review included advice that stated "the classification of bioregions into ecological units should strive to incorporate detailed knowledge of species distribution and abundance patterns as well as their interactions with their habitat and other species" and suggested that when biological data are not available, geophysical and oceanographic factors can be used when there is evidence that these factors can discriminate among habitat and community types (DFO 2012b), The science advice also stressed that classification below a scale for which not enough data are available to create an accurate classification should be avoided. The scale of available data is a key limiting factor for the PMECS level that can be populated, and here we are restricted to data sources that have reasonably complete and systematic coverage at the scale of interest. In this study, we concentrate on developing methodologies that will classify areas below the bioregion level (see Table 4). We attempt to classify large-scale patterns in biodiversity and their associations with large-scale variability in the environment.

Our specific objectives to move PMECS from a conceptual framework to implementation are:

1. Develop a central geodatabase by compiling existing spatially referenced biotic and abiotic data for used in marine spatial planning in Canada's Pacific Region;
2. After examination of available data sources, develop analytical methodologies to progress PMECS towards implementation in the Northern Shelf and Southern Shelf Bioregions (NSB, SSB; Table 4);
3. Apply classification methods to Northern and Southern Shelf Bioregions, with a focus on subtidal areas, to populate into Level 4 of the PMECS framework (Biophysical units);
4. Identify species commonly associated with each of the ecological units developed for Level 4a and b (Biophysical units)
5. Apply classification methods to NSB/SSB to populate Level 5 (Geomorphic units) of the PMECS framework.
6. Propose quantitative methods suitable for classifying areas at finer scales (lower levels) of the PMECS hierarchical framework

1.6 APPROACH

There have been several literature reviews on marine ecological classifications (see Lund and Wilbur 2007, Greene et al. 2008, Harris 2012a), including the review done by Robinson et al. (2015) to build the conceptual PMECS framework. We followed the recommendations of Robinson et al. (2015) and the prototype PMECS shown in the DFO (2013) for guidance in developing the methodology to apply PMECS to the NSB, although some modifications were needed to implement the system (see next section). Given that the conceptual PMECS framework followed the Last et al. (2010) approach, we use the methodologies used in the application of Last et al. (2010) on Australia's North West Shelf (Lyne et al. 2006) as guidance to populate the PMECS layers in NSB and SSB. To avoid redundancy of ongoing work on benthic classifications in the region, we reviewed and integrated existing and ongoing classification methodologies where appropriate (e.g., Gregr et al. 2013). Table 4 shows the classification structure of PMECS that we developed to build upon the conceptual framework of Robinson et al. (2015), outlining data needs and analytical methods to populate each level of the hierarchical classification system.

1.6.1 Modifications to conceptual framework to move towards PMECS implementation

To move the conceptual or prototype PMECS developed by Robinson et al. (2015;Table 2) towards implementation, we made some minor modifications to the framework. The modifications were made primarily based on the types of data available and the management objectives for MPA network planning. Upon review of the benthic column of Table 2 and Table 4 there are four main differences:

1. We renumbered the levels within the framework to begin with Level 1 rather than Level 0 to be consistent with other classification systems.
2. We used the Zoogeographic Provinces to define the "Province" level (Level 2, Table 4) for the benthic classification (Table 2 only uses it for the Pelagic example), because demersal fish species were included in the analyses to delineate those provinces (Allen and Smith 1998) and incorporating the biogeography strengthens the biological foundation of the classification system.
3. We removed the BCMEC Ecosections (Level 3, Table 2) from the benthic classification for two reasons:
    a. The spatial extent of Ecosections and Bathomes (Level 4, Table 2) were similar so the resulting classification would not be spatially hierarchical; and
    b. We tested the biological relevance of the Ecosections to benthic biodiversity and found that only the Continental Slope Ecosection unit showed strong evidence of a distinct biotic assemblage (See Appendix A).
4. We developed new nomenclature for the levels of PMECS in order to develop a harmonized classification system with other regions in Canada (i.e., Maritimes region, Greenlaw et al 2015; see Table 4).

Table 4. Levels, definitions, data needs, outputs and classification methods of the hierarchical Pacific Marine Ecological Classification System framework. The italics indicate new analytical tools developed in this paper to operationalize the framework definition at that level. The remaining levels have been developed previously or are currently being developed and applied by other scientists in the region. All levels presented here are suitable for informing ecosystem-based management decisions. "Species-based management levels" of the framework (species, populations and genes, Levels 10–12, Table 2) are not included in our revised table. All levels in this table refer to a ***benthic*** *ecological classification, a pelagic system needs to be developed separately (Robinson et al. 2015).*

PMECS Level	Spatial extent	Framework definition	Data used to generate/populate level	Output Unit in Pacific Region	Method used or Reference
1. Realms	>10000 km	Ocean basin	Oceanography, large-scale species distributions, and expert opinion.	Temperate North Pacific	Spalding et al. (2007), UNESCO (2009)
2. Province	1000s of km	Based on large-scale patterns of endemism	Biogeographical data on fish (two distinct fish fauna meet on BC coast), algae, and intertidal invertebrates.	Oregonian (also referred to as Columbian) & Aleutian transition zone; Biogeographic break near Brooks Peninsula for algae on west coast of Vancouver Island but closer to Cape Flattery in WA state for fish fauna.	Allen and Smith (1988), Druehl (2000), Fenberg et al. (2015), Robinson et al. (2015)
3. Bioregion	1000s of km	Distinctive recurring and smaller-scale physical oceanographic processes (e.g., separation between California current and Alaska current regions)	Oceanographic processes and bathymetry.	Northern Shelf, Offshore Pacific, Southern Shelf and Strait of Georgia	DFO (2009)
4. Biophysical Units	*100s – 1000s of km*	*Distinct physiographic and oceanographic conditions/processes, including bathymetry, related to biotic composition if data are available or evidence in the literature.*	*Benthic species composition, bathymetry, temperature (bottom and sea surface), salinity and currents.*	*Dogfish Bank, Other Banks, Shelf, Troughs, Slope*	*Species composition cluster analysis, and multivariate analysis of environmental data to delineate biogeographical units (see methodology next section)*
5. Geomorphic Units	*100s of km*	*Discrete geomorphological structures assumed to have distinctive biological*	*High resolution bathymetry (≤ 75m) and derived metrics such as Benthic*	*Ridge, Canyon, Gentle Slope, Steep Slope*	*Methodology described here; based on fine-scale and broad-scale*

PMECS Level	Spatial extent	Framework definition	Data used to generate/populate level	Output Unit in Pacific Region	Method used or Reference
		assemblages; Individually defined by shape, size and topographic variation. May span other levels of hierarchy.	*Positioning Index (BPI).*		*BPI and other metrics derived from bathymetry to delineate geomorphological features on the seabed (following Manson (2009).*
6. Biotopes (Habitats and communities)	100s of m – 100s of km	Discrete taxonomic assemblages characterized by associated substrate and environmental factors.	High resolution bathymetry (75 m), 6 sources of bottom type data (ShoreZone, 3 observational and 2 grab sampling datasets containing substrate data, and tidal energy). Additional data sources will be incorporated when available including analyses of backscatter data, more acoustic surveys, multibeam sonar and all available biotic data to develop community-substrate type associations.	Hard, Mixed, and Soft substrate types. These major abiotic units act as surrogates for assemblages living on or in these substrates.	Bottom Patch Model described in Gregr et al. (2013) and Gregr et al in prep.
7. Biological Facies	100s m	Groups of biogenic or foundation species identified by one or more indicator species. BFs are patchy and nested within biotopes. Most examples are biogenic habitats.	Direct survey data mapped (underwater imagery, benthic samples, etc.).	E.g., Sponge reefs, kelp forests, eelgrass beds	e.g., Delineating and mapping sponge reefs: (Conway et al. 1991, Conway et al. 2005, Conway et al. 2007)
8. Micro-assemblages	1 cm – 10s m	Small scale assemblages of often highly specialized species. Will be associated with biotope but may not or may not be associated with a biological facies.	Direct survey data, field surveys, benthic samples.	E.g., kelp holdfast assemblages	Additional research is required to identify the suite of micro-assemblages in the Pacific Region (Robinson et al. 2015)

2 DATA SOURCES

The initial step in developing PMECS was to compile existing and relevant marine abiotic and biotic spatial data from Canada's Pacific Region into a central geodatabase and evaluate their suitability for use in the ecological classification analysis. This step included liaising with contacts from DFO, BC Ministry of the Environment (MoE), BC Ministry of Forest, Lands and Natural Resource Operations (FLNRO), Natural Resources Canada (NRCan), Environment Canada (EC), Parks Canada (PC), First Nations (FN), Canadian Hydrographic Service (CHS) and natural history museums (Royal BC Museum, UBC Beaty Biodiversity Museum). We also engaged non-profit organizations that have compiled and provided freely available spatial data, including BCMCA (British Columbia Marine Conservation Analysis) and OBIS (Ocean Biogeographic Information System).

2.1 DEVELOPMENT OF THE PMECS GEOSPATIAL DATABASE

Although the initial application of PMECS is to benthic ecosystems in the NSB and SSB, we collated benthic and pelagic species data from all areas within Canada's four Pacific bioregions, out to the Exclusive Economic Zone, with the intent to store the data for future applications of PMECS and other marine spatial planning analyses. All spatial data acquired for the project have been annotated and are currently housed on a DFO server. All data layers that were collected or created are stored using a systematic filing system, with detailed metadata associated with each layer that include source data, references and any additional analyses that were conducted to generate the layer. All spatial data that were compiled for PMECS are housed in this database, and although not all of the data are suitable for use in the classification algorithms, they can be used for other objectives related to marine protected area network design and marine spatial planning in general. Currently, the PMECS geospatial database has about 600 abiotic and biotic spatial layers.

The development of the geospatial database is a valuable output of this project and can be accessed by contacting the authors of this document. It can be used for various marine spatial planning projects including identifying ecologically and biologically significant areas, mapping the distributions of species at risk and their critical habitat, developing species distribution models, modelling species responses to climate change scenarios, and a broad range of other analyses related to marine conservation. Datasets included in this database were first entered in their original form (i.e., unedited, uncorrected); any modified versions of the data were also included after quality-control or analytical processing (see sections 2.1.2 and 2.1.3). The types of data held in the PMECS database are described below.

2.2 BIOTIC DATA

Biotic data were compiled from a number of sources, including the online data repository OBIS, the Royal BC Museum Invertebrates (RBCM-I) and fish (RBCM-F) collections, and from DFO Shellfish Commercial Log (SF Log), Shellfish Research Biology (SF Bio), and Groundfish (GF) data holdings. Data sources were initially considered for inclusion in the analysis if they contained records having, at a minimum, a species name and a point location (i.e., a latitude and longitude). Range maps (i.e., polygons) generally overestimate the true distribution of species occurrence, because species do not occur in every location within these ranges (Hurlbert and Jetz 2007). Specifically, it is difficult to assess the degrees to which species co-occur in an assemblage or the degree to which species distributions correlate with environmental variables based on range information alone. Therefore, data presented as polygons were not considered for the current analysis, but are nevertheless stored in the geospatial database. The following data sources are present in the PMECS geospatial database but only a subset (see section 3.2.1) was included in the final analysis.

2.2.1 OBIS

OBIS is an open-source online repository that contains species locality point data contributed by a range of researchers, organizations, museums, and other collection managers. The OBIS database was queried on 26 August 2014; all data sources within OBIS that included species locality points in the marine regions of BC were selected and exported. The OBIS dataset contains approximately 40,000 marine, species-level records (~130,000 records including higher taxa) from 15 sources, including museums, DFO, Environment Canada/Canadian Wildlife Service, citizen science initiatives, graduate theses, and other published sources. We did not collate the citizen science records for this analysis due to concerns about data quality.

2.2.2 Royal BC Museum (RBCM)

Two datasets were obtained from the Royal BC Museum, an invertebrate collection dataset (RBCM-I, includes all records in the database as of 9 September 2014) and a fish collection dataset (RBCM-F, obtained 14 October 2014). The RBCM-I dataset contains approximately 33,000 marine, species-level records (~63,000 records including higher taxa), while the RBCM-F dataset contains approximately 5500 marine, species-level records (~15,000 records including higher taxa).

2.2.3 Shellfish Data Holdings: SF Log and SF Bio

DFO Shellfish Commercial Log (SF Log) and Research Biology (SF Bio) data holdings, which contain databases for each shellfish fishery or research program, were queried between October and December 2014. The commercial log databases contain records of commercial fishery catch as reported to DFO, and primarily consist of records of single target species (Table B 1). Exceptions are the Scallop by Trawl, Shrimp by Trawl, and Tanner Crab databases, which contain records of multiple non-target species. The research biology databases contain records of research surveys undertaken by DFO shellfish biologists under specific programs, many of which include records of target and non-target species.

For most databases, georeferenced records from the year 2000 and later were extracted; actual years extracted varied due to differences in data availability or fishery openings. Additional limits were imposed to ensure reliability of extracted records, including an acceptable "usability" code (usability = 0, assigned by DFO to indicate records that had no quality control flags in the database, e.g., caused by incomplete trawls or errors during collection). Where multiple locations were reported for a record (e.g., start and stop locations for a trawl), the average of available latitudes and longitudes was calculated. Although fishing events often do not occur in a straight line (e.g., when following a bathymetric contour), any spatial errors caused by averaging were accounted for when the resulting data points were aggregated into 4 km grid cells (see section 3.2.2)

2.2.4 Groundfish Biological Survey Database (GF Bio)

Groundfish biological survey data were obtained from the DFO GFBio database. This dataset included 136,721 records collected between 2003 and 2013 from 16 survey activities (Table B 3). Effort metrics were provided depending on the type of gear used in each survey, including number of hooks set and retrieved (longline surveys), number of traps (trap surveys), and time duration of trawl (bottom and midwater trawls). Catch was reported as either weight of catch in kg or number of individuals; records that contained either a weight or a count (including records of “trace” material”) were considered “present” for use in the present-absence analysis (see section 3.2.2). A summary of the general survey design for each survey activity is available in Appendix B.

2.2.5 Biotic data preparation

The biotic data were quality-checked and pre-processed before being saved into the PMECS geospatial database. All taxonomic names associated with records were compared to the World Register of Marine Species (WoRMS) or the Integrated Taxonomic Information System (ITIS) to check for synonyms and misspelled names in the datasets. The WoRMS database was first queried using an automated R tool, which resolved most of the synonyms, but could not process misspellings. The remaining records were manually processed. Records with taxonomic names that were not present in WoRMS or ITIS were excluded. Although DFO maintains a regional species name table for its databases that is generally consistent with the WoRMS and ITIS classifications, we found a number of differences in “accepted” scientific names due to recent taxonomic revisions on the online sources, differences in standard use of subspecies/subgenera, and spelling mistakes in the DFO table. The WoRMS and ITIS names were chosen as the standard for this work to facilitate the potential integration with non-DFO datasets; there is no methodological or analytical concern using one set of standard names over another.

Records that were not georeferenced, or which were referenced to an “approximate latitude/longitude” were excluded. Records of terrestrial or freshwater species, birds, mammals, insects, arachnids, diatoms, fungi, and records that were otherwise of questionable quality were also excluded. Examples of these exclusions include records with ambiguous names, such as names shared between plant and animal species with no higher taxa given, or records with georeferencing that placed species in questionable areas (such as intertidal barnacles in the middle of the ocean).

The higher-level taxonomy for each record was appended to the record. Species that were identified in the initial dataset as “cf” or “nr” or “?” (e.g., *Suberites cf. affinis*, *Suberites bursa*?) were changed to genus-level records. For this analysis, these genus-level records were excluded and only species-level records were used. Analysis of communities using higher-level taxonomic groups in place of species may be appropriate and informative in some cases (Włodarska-Kowalczuk and Kędra 2007, Bett and Narayanaswamy 2014), but mixing taxonomic levels in a single analysis (e.g., having some groups identified to species and others to order) can obscure community patterns (Hernandez et al. 2013).

Habitats were assigned using existing information about habitats used by each species, as designated by the literature, field guides or online databases such as SeaLifeBase or FishBase. These online databases provided information on most taxa, classifying them as benthic, pelagic, terrestrial, freshwater, and combinations thereof (benthopelagic, freshwater/marine, etc.). A “reef-associated” category was also provided. A number of taxa could not be and were not classified, including mostly meio/macro/infaunal groups (e.g., Annelida, Nemertea, Branchiopoda, Platyhelminthes). Only a subset of all biotic data that we compiled was used in the present PMECS analysis, as described in section 3.2.1.

2.3 ABIOTIC DATA

We collated abiotic (environmental) data layers from multiple sources including DFO, U.S. National Aeronautics and Space Administration (NASA), Canadian Hydrographic Service (CHS), BCMCA, and the U.S. National Oceanic and Atmospheric Administration (NOAA). We compiled 59 abiotic raster layers for the geospatial database (Table 5). Many of these data are derived or modeled and it is important to note that there are inherent errors associated with these kinds of data layers. Although quantifying these errors and their impacts on associated analyses is beyond the scope of this project, with the description of each layer below, we state known uncertainties or possible sources of error.

2.3.1 Depth and derivatives

A 100 m bathymetry raster, derived from a Canadian Hydrographic Service (CHS) layer and available from BCMCA was used as our depth layer. Slope, aspect, and rugosity (surface area to planar area) layers were calculated from the depth layer using the Benthic Terrain Modeller extension for ArcMap. Depth data were originally derived from CHS chart contours, interpolated to a raster, and combined with other bathymetry layers to obtain a continuous surface throughout Canada's Pacific Region (i.e., within the Exclusive Economic Zone). There are increased artifacts and decreased accuracy at deeper depths (Gregr 2012), however the metadata for the depth layer (Gregr 2012) note that the errors are "self-consistent" and suggest that the layer is suitable for coast-wide or regional analyses. Slope and rugosity calculations from this layer will also include errors associated with the original raster.

2.3.2 Chlorophyll A

Global sea surface chlorophyll-A concentrations (mg m^{-3}) were obtained from 4 km resolution, level 3 AquaMODIS data (NASA Ocean Color). Monthly and annual composite products were downloaded for the period from January 2005 to December 2014. Monthly files were binned into three-month long seasons following NASA (Help Page for Level-3 Browse Interface) and Foreman et al. (2008): winter is January to March, spring is April to June, summer is July to September, and fall is October to December. Annual minimum, maximum, mean, and range layers were created from the 10 annual composite layers and from the 30 layers in each season, for a total of 20 chlorophyll input files

Remotely sensed data, in general, often have gaps due to clouds, aerosols, and sun glint (Gregg and Casey 2007), which can introduce errors or biases when averaged or aggregated (Tyberghein et al. 2012). Extrapolation towards coastal regions can also produce errors (Tyberghein et al. 2012); NASA's algorithms for the level 3 layers do not process sites near to shore (i.e., raster cells that intersect land), so most of the areas near the coast do not have associated chlorophyll data. AquaMODIS tends to overestimate chlorophyll A concentration in the North Pacific Ocean due to lost sampling days from cloud obscuration (Gregg and Casey 2007). However, this effect is consistent over years and within basins (Gregg and Casey 2007) and therefore presumably consistent over our study area and so should not significantly affect our analysis.

2.3.3 Sea surface temperature

Global sea surface temperature (SST) data (4 km resolution, level 3 AquaMODIS 11 µm daytime SST) were also obtained from the NASA Ocean Color website. Seasonal (winter 2002–2013; spring, summer, and fall 2002–2014) and mission-long ("overall"; July 2002–January 2015) composites were downloaded and used without transformation, for a total of 5 SST input files. SST data layers are subject to similar remote sensing errors as outlined above for Chlorophyll A.

2.3.4 Bottom Salinity, temperature, currents

Bottom temperature, salinity, and tidal and non-tidal flow values were derived from an ocean circulation model (Foreman et al. 2008). Seasonal, maximum, minimum, and range layers were created for bottom temperature and salinity (date range 2003–2009). Flow (non-tidal current velocity) was represented by overall, north-south, and east-west values for summer and winter. One layer each represented tidal speed and direction, as these values are non-seasonal.

The distance between nodes of the ocean circulation model varies from 100 m in some small coastal channels to more than 50 km in deep, offshore areas. These node points were converted to shapefiles and interpolated to a 500 m raster using "spline with barriers" in ArcMap 10.1.

There are several sources of error in the underlying numerical model used to estimate ocean circulation and associated attributes, including boundary conditions (tidal amplitude and phase provided by a coarser model), interactions between tidal components, the climatology used, bathymetry, and the representation of important bathymetry and coastline features (Pramod Thupaki, DFO, Institute of Ocean Sciences, Sidney, B.C., personal communication, 2015). Errors are pronounced for this particular dataset along the coast and in the coastal fjords. A more recent model built using a higher resolution grid has been shown to perform better, especially in long narrow fjords/channels (P. Thupaki in prep.), but was not available for our entire study area.

2.3.5 Habitat-Template model

Layers for a Habitat-Template model for the Pacific continental shelf were provided by Ed Gregr (Gregr et al. in press[1]), including adversity and disturbance variables. Adversity is defined as the severity of the environment, and incorporates temperature range, mean summer temperatures, and food availability (spring chlorophyll-A bloom frequency, summer stratification index, and depth). Disturbance is defined as the potential for physical disturbance due to mobilization of sediment by water flow, and incorporates bottom friction velocity (tidal, ocean current, and wave shear stresses) and critical shear velocity (critical shear stress, standard gravity, grain size, grain density, and viscosity). For more information about this layer and how it was generated see Gregr et al. (in press) and for a more general reference, Kostylev et al. (2005).

2.3.6 Grain size model

Ed Gregr provided his grain size model, part of the Habitat-Template model (Gregr et al. in press[1]). The grain size model is a statistical model of benthic particle size, derived from NRCan Expedition data, incorporating depth, slope, bottom current, BPI and predicted grain size using generalized additive models. These layers do not extend onto the continental slope. The details of this model are summarized in Gregr et al[1].

2.3.7 Nutrients

We obtained surface nitrate, silicate, phosphate, pH, and dissolved oxygen values from Bio-ORACLE (Tyberghein et al. 2012), an open-access set of global rasters at 9.2 km resolution. These data are derived from NOAA's World Ocean Database (WOD) 2009, with original *in situ* data sources ranging from the late 19th to late 20th century. Data flagged as erroneous in WOD were removed by the Bio-ORACLE creators, and a data quality assessment showed that the highest uncertainty was at latitudes > 70 ºN or ºS and in some unsampled areas in the open ocean (Tyberghein et al. 2012). The major source of uncertainty using this layer in the model presented in this paper is the mismatch in scale between the nutrient layer (9.2 km) and the conditions felt by individual organisms (a common issue for all species distribution modeling exercises).

---

[1] Gregr, E.J., Gryba, R., Li, M., and Hannah, C.G. in press. A benthic habitat template for Pacific Canada's continental shelf. Canadian Technical Report of Hydrography and Ocean Sciences

Table 5. List of environmental data layers compiled and/or generated for use in environmental analysis

Feature	Derivatives	Source	Native Resolution	Number of layers
Chlorophyll A	Annual and seasonal (3-month) minimum, maximum, mean, and range	NASA (AquaMODIS)	4 km	20
Sea Surface Temperature (SST)	Seasonal (3-month) and 13-year ("overall") composites	NASA (AquaMODIS)	4 km	5
Depth	Depth, aspect, slope, rugosity (surface to planar area)	BCMCA/CHS	100 m	4
Bottom temperature	Spring, summer, winter, fall, max, min, range	Mike Foreman, DFO	500 m raster; initial node spacing varies from 100 m nearshore to > 50 km offshore	7
Bottom salinity	Spring, summer, winter, fall, max, min, range	Mike Foreman, DFO	500 m raster; initial node spacing varies from 100 m nearshore to > 50 km offshore)	7
Flow (non-tidal current velocity)	Summer (overall, north-south, east-west) and winter (overall, north-south, east-west)	Mike Foreman, DFO	500 m raster; initial node spacing varies from 100 m nearshore to > 50 km offshore	6
Tidal speed and direction	Bottom RMS speed over 29 days of the M2 component	Mike Foreman, DFO	500 m raster; initial node spacing varies from 100 m nearshore to > 50 km offshore	2
Grain Size	-	Gregr et al.[1]	100 m	1
Adversity	-	Gregr et al.[1]	100 m	1
Disturbance	-	Gregr et al.[1]	100 m	1
Nitrate	-	Bio-ORACLE	9.2 km	1
Silicate	-	Bio-ORACLE	9.2 km	1
Phosphate	-	Bio-ORACLE	9.2 km	1
pH	-	Bio-ORACLE	9.2 km	1
Dissolved oxygen	-	Bio-ORACLE	9.2 km	1
Total	-	-	-	**59**

3 LEVEL 4A AND B: BIOPHYSICAL UNITS

3.1 APPROACH

We developed a methodology to classify Level 4 of PMECS to differentiate major biological assemblages delineated using abiotic variables at a spatial extent of 1000s of km (Table 4). This level of the PMECS framework is analogous to the levels described as "bathomes" in Last et al. (2010) and Robinson et al. (2015). Last et al. (2010) define bathomes according to the depth distribution of biota, and describe some of the limiting factors of this layer as physiological constraints on species depth distributions and depth-related differentiation in habitat distribution. To operationalize this level, one option is to divide the coast into pre-defined depth zones based on contours that are likely to have some association with the distribution of biota (i.e., coastal 0–50 m, shelf 50–200 m, slope >200 m). However, a more biologically relevant approach, and the one chosen here, is to examine the distribution of species across variable depths and allow the biological patterns to inform the breaks in species assemblages across the study area, and then correlate these assemblage patterns with environment variables to determine whether these patterns are predictable.

In an application of the Last et al. (2010) classification framework to characterize the ecosystems of Australia's North West Shelf, Lyne et al. (2006) use a cluster analysis of fish species to better understand the depth distribution of species in their study area. In this study, we use similar methodologies to Lyne et al. (2006). Cluster analyses are commonly applied in community ecology to investigate spatial patterns of species composition and when combined with multivariate environmental analyses can be used to understand biogeographic and ecological patterns across space (Kreft and Jetz 2010). Several studies have used this approach in the Pacific Ocean, including an ecological classification of fish habitat in California (Allen and Pondella 2006), an analysis of the biogeographic structure of the northeastern Pacific rocky intertidal zone (Fenberg et al. 2015) and a classification of groundfish species assemblages in Hecate Strait, BC (Fargo 2012).

Here, we use a cluster analysis based on the similarity of species composition among sites to better understand patterns in biological diversity in our study area. We also use an indicator species analysis to identify which species are most commonly associated with each cluster (Dufrêne and Legendre 1997). This analysis allows us to have a list of associated species that are likely to be present within each ecological unit. Upon completion of the cluster analysis, we conduct an environmental analysis of those clusters using a random forest approach (Breiman 2001) to determine if the environment can explain the differences in species assemblages in our study area. A strength of using the random forest approach is that as a predictive model, we can use the relationships between the environment and the presence of the biological assemblages to predict which assemblage (identified by the cluster analysis) is likely to occur in areas where no biological information exists. The predictive performance of random forest models, which are regularly used in species distribution modelling studies, is typically equivalent to or better than other statistical and machine-learning methods in comparison relating ecological and covariate data (Prasad et al. 2006, Cutler et al. 2007). Fenberg et al. (2015) used random forest to examine the biogeographic structure of the northeastern Pacific rocky intertidal zone, including the BC coast, and found that sea surface temperature (SST) and nutrients were an important predictor of biogeographic structure in the area. A key feature of random forest modeling is that as a classification/regression tree method, nonlinearities and interactions among predictor variables are considered by default and it can cope with highly correlated predictor variables (Strobl et al. 2008).

The vast majority of benthic marine organisms are limited by some combination of depth, substrate type, temperature and salinity, but the complexity of the relationships to these variables are less well understood (Roff and Zacharias 2011). A review of 57 studies on mapping marine benthic communities found that water depth was the most useful surrogate for delineating benthic communities followed by substrate type (Harris and Baker 2012a). Unfortunately, a reliable substrate type layer at the scale needed for this analysis does not exist; therefore, we predicted that broad-scale biological assemblages in our study area will be mainly driven by changes in bathymetry.

3.2 METHODOLOGY AND APPLICATION

3.2.1 Selection of data

Prior to completing the analysis to delineate the Biophysical unit, we completed some initial exploration of the biological data collated in the PMECS geospatial database to determine which datasets were appropriate for use in the development of an ecological classification. Given the recommendation to produce a classification with a strong biological component, and the fact that a physiographic classification exists in BC (BCMEC), we aimed to maximize the use of biological data while limiting errors associated with using existing datasets from multiple sources. To do this, we developed conservative criteria for the inclusion of biological data. As a first pass, data containing no species-level information (e.g., only recorded as genus, phylum, or order) were not considered, and because of difficulties confirming and standardizing the names of marine algae in our database, marine plants and algae have not been incorporated into the analysis at this time.

Our approach to populating Level 4 of PMECS has two steps. The first is to examine biological patterns using a cluster analysis of species composition, where sites with similar species are grouped together into distinct biological assemblages. The second step is to use a random forest model to identify environmental correlates of the identified biological assemblages, and use this model to predict the biological assemblage present in areas with no biological data. The underlying assumptions of our approach are that the biological data are representative and independent, and that any environment-assemblage relationships calculated by the random forest model hold when predicting assemblages in areas within the study area that lack biological data. The latter point is addressed, in part, by running the model on training and testing subsets of the data and examining model performance metrics.

To best adhere to assumptions that our data are representative and independent, and to reduce sampling bias, we only selected data from the PMECS geospatial dataset that met the following criteria:

1. Multispecies surveys with no limit on the number of species recorded; and
2. North – South spatial coverage across the study area (localized or patchy data sources were excluded, unless that dataset covered sites that would otherwise not be represented in the analysis).

We also considered the uncertainty and inherent error in some of the abiotic data layers, particularly in narrow channels, fjords and other coastal areas (as outlined in Section 2.3) when finalizing spatial extent of our study. Because of the lack of (or error-prone) abiotic data in sites near the coast, the complexity of the BC coast, and local processes and environmental drivers unique to the coast (such as freshwater and nutrient inputs from land, and localized currents due to topography of islands/peninsulas/fjords etc.), we removed all sites that intersected with land from the classification analysis. Due to their complexity and uniqueness, coastal areas and the Strait of Georgia Bioregion need to be examined separately and also were removed from our analysis.

After filtering data sources through these criteria, our final dataset included data from the Groundfish research surveys (GF Bio; 2003–2013), and two datasets from the Shellfish research data holdings (SF Bio): the Tanner Crab surveys (2000–2006) and the Crab Trap surveys (2000–2014). The Tanner Crab data was included because of its coverage of the continental slope, which the otherwise extensive Groundfish surveys did not cover. The Crab Trap surveys were included to fill in data gaps on Dogfish Bank (off northeast Haida Gwaii) despite being relatively localized and including some single-species records (see discussion in Appendix B). Two other potentially appropriate datasets were not used in an effort to be as conservative as possible in data inclusion. Data from commercial groundfish fisheries were excluded because of the lack of gear type standardization, and the Shrimp Trawl research survey (Multispecies Small Mesh Trawl survey) data was excluded because it had a localized/patchy extent that was sampled with a different gear type than the groundfish trawls, potentially introducing spatial biases in the analysis.

We selected records of “bottom dwelling” (habitats assigned as benthic, demersal, benthopelagic, or reef-associated) fish and sessile or moderately mobile invertebrate species from these datasets to classify benthos in line with the goal of the Biophysical unit analysis. Plankton, meiofauna, infauna, diatoms, algae, plants, etc. were excluded. Four hundred and eighty-eight bottom-dwelling species were present within the study area and are listed in Tables D 1 and D 2 of Appendix D. Taxa represented in this dataset were Pisces, Arthropoda (Cirripedia, Decapoda), Tunicata, Anthozoa, Echinodermata (Asteroidea, Crinoidea, Echinoidea, Holothuroidea, Ophiuroidea), and Mollusca (Bivalvia, Cephalopoda, Gastropoda, and Polyplacophora). All sponges and many coral species were excluded from the present Biophysical unit analysis because of lack of species-level records (see Section 2.2). However, habitat-forming corals and sponges are represented at lower levels of the PMECS hierarchy (Table 4) as “Biological Facies” so these taxa are included in the classification at finer scales. For the cluster analysis, fish and invertebrates were considered together in one analysis as we found little biological justification for conducting separate analyses of these taxa. Preliminary sensitivity analyses showed that cluster analyses based on fish alone produced similar large-sale patterns (e.g., showing slope, troughs, and Dogfish Bank; see results in section 3.3) as the analysis based on invertebrates alone; combining the taxa increased the number of records and the number of sites included in the cluster analysis, improving the resolution of the final output.

3.2.2 Cluster Analysis Methodology

Cluster analysis was used to assess patterns of species assemblages. The output of a cluster analysis is a dendrogram (tree or cluster diagram) that displays hierarchical groupings (clusters) of sites based on the similarity of their species assemblages (Legendre and Legendre 2012). Cutting the tree at a given distance produces discrete clusters, for which the assemblages can be identified and analyzed. The closer a cut is to the branches of the dendrogram, the progressively more similar the species assemblages are within the resulting clusters. When the sites within each of these clusters are plotted on a map, spatial patterns of species assemblages can be examined (Kreft and Jetz 2010). Cluster analyses do not involve any *a priori* assumptions about how or why species are distributed among sites; they are a descriptive analysis that show patterns of species assemblages (Legendre and Legendre 2012). The only assumption of a cluster analysis (using “average” or unweighted pair group method with arithmetic mean, “UPGMA”; see following paragraph) is that the input sample data are representative of the larger population (Legendre and Legendre 2012).

A standard methodology was used to differentiate major biological assemblages within NSB and SSB. Sampling units were defined as 4 km grid cells ("sites") and all species that were found within that grid cell were considered to be located at the same site. This cell size was chosen after exploratory analyses showed that larger cells (e.g., 8 km) were too coarse, while smaller cells (e.g., 2 km) generally captured only one or a few sampling events, limiting the number of species at each site (see footnote 3 on next page) and giving a surface with many gaps that limited interpretation. Data were converted to presence-absence of each species at each site. To calculate the similarity of species composition at each site relative to all other sites, we calculated a matrix of pairwise β_{sim} distance values. The β_{sim} distance (also called Simpson distance or Simpson dissimilarity index; Koleff et al. 2003, Baselga 2010). The Simpson dissimilarity index[2] is a richness-independent dissimilarity measure that isolates the "turnover" (as opposed to nestedness) component of beta-diversity (Baselga 2010). We selected this index over richness-dependent indices, such as Bray-Curtis, to reduce the influence of unequal sampling across our study area. This metric has been shown to perform well for presence-absence data (Koleff et al. 2003). β_{sim} is defined as:

$$\beta_{sim} = 1 - a/(\min(b,c) + a)$$

where *a* is the number of shared species between sites, and *b* and *c* are the respective number of species unique to each site. β_{sim} values range from 0 to 1, with 0 indicating no dissimilarity (full similarity, identical species composition) and 1 indicating full dissimilarity (no similarity, no shared species; Kreft and Jetz 2010). The matrix of pairwise β_{sim} values was used to create a dendrogram using average or unweighted pair group method with arithmetic mean, UPGMA, clustering. To assess the performance of the β_{sim} distance compared to other (dis)similarity measures, we compared the cophenetic correlation coefficients (a measure of goodness of fit) for the dendrograms produced using the β_{sim}, Sorensen, Jaccard, and Ochiai distances. β_{sim} was found to perform almost twice as well as the other distances (β_{sim} cophenetic $r = 0.672$, vs. 0.356, 0.337, and 0.365 for Sorensen, Jaccard, and Ochiai, respectively), supporting its suitability for the present analyses.

Species with low frequency in the dataset (i.e., not found at many sample sites) are routinely removed from community cluster analyses because they can add noise to multivariate analyses and provide little information in addition to that obtained from more common species (Gauch 1982, McCune and Grace 2002), but the legitimacy of removing these species depends upon the objective of the study (Poos and Jackson 2012). Here, our objective is to examine broad-scale patterns in biological communities that are somewhat stable over time and that are representative of the diversity of the area, so basing this analysis on more common species is in line with our objective. A species may be infrequently recorded in a given dataset for several reasons, including: the species is truly rare in the study area; the species is present but not detected in the current sample (sample bias); or the species is present but identified inconsistently (observer bias). Choosing a threshold to exclude infrequent species is somewhat arbitrary, but literature-reported thresholds for exclusion include species found in less than 5% of sites (McGarigal et al. 2000), 10% of sites (Marchant 1990, McCune and Grace 2002) or even higher (Marchant et al. 1997). To maximize the inclusion of species while also reducing noise and potential biases in the analyses, we chose a conservative exclusion threshold and removed species reported in less than 1% of sites.

---

[2] Note that the subscript "sim" stands for Simpson, not similarity. β_{sim} is a dissimilarity metric.

Sites where only one species was recorded were also removed. The inclusion of sites with only one species can artificially impact β_{sim}, which in turn constrains the topology of the output dendrogram and obscures patterns of true similarity among more data-populated sites[3]. Our final dataset included 174 species (96 species of demersal fish and 78 species of benthic invertebrate) and 3615 sample sites (Figure 3, see Appendix D for included and excluded species).

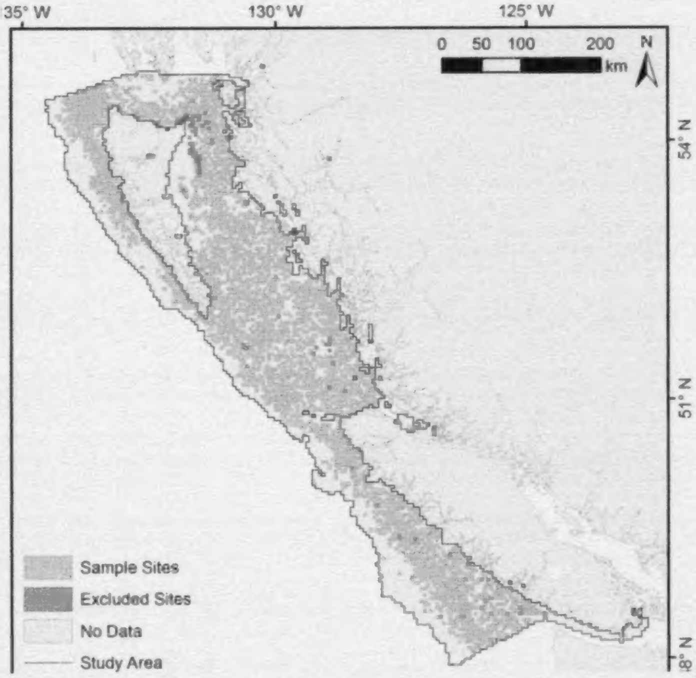


Figure 3. Map of study area and sample sites. The study area includes 6875 sites (4 km grid cells) within the Northern Shelf and Southern Shelf Bioregions (total area is 110,000 km^2). Excluded sites (purple) were not considered for the cluster analysis because they only included 1 species, and gray areas within the study area had no biological data; however, these sites are classified in the following random forest analysis (section 3.2.4). Sites with biological data that were included in the cluster analysis are shown in orange (n = 3615).

3.2.2.1 Choice of dendrogram cutoff

Determining the appropriate number of clusters (*k*) is an enduring issue in the cluster analysis literature (Milligan and Cooper 1985). In other biogeographic studies, several different types of stopping rules have been used, for example, a minimum number of grid cells (sites) per cluster (Williams et al. 1999), a predetermined level of dissimilarity (Proches 2005), or the height of the nodes of dendrogram and various metrics of relative endemism within clusters (Kreft and Jetz 2010). Given that the objective our study was to delineate the study area into biologically relevant ecological units to fulfil the "representativity" criterion for MPA network planning, we wanted to maximize the number of clusters to ensure all ecosystems at this scale were captured while also maximizing the number of sites classified into geographically coherent clusters (thereby minimizing unclassified sites). To determine the optimal cut-off we examined three metrics:

---

[3] When comparing a site with only 1 species to any other site, the maximum number of shared species is 1, meaning the only potential β_{sim} values are 0.5 [1 – (1/1+1)], 1 [1– (0/1+0)], or 0 [1-(1/0+0)]. This limited range of possible β_{sim} values constrains the potential topology of the output dendrogram, preventing meaningful results.

1. The proportion of sites in the most populated clusters;
2. The spatial coherence (clumping) of the sites in each cluster; and
3. The variance in number of sites per cluster.

We examined two cut-off points resulting in broad scale "Level 4a" and finer-scale "Level 4b" Biophysical clusters that were both analyzed further in the same way. Figure 4 shows how the number of sites in the 8 most populated clusters increases with increasing β_{sim} distance, while Figure 5 shows that the variance in cluster size increases with the β_{sim} distance and the number of sites included in the top clusters. The cut-off was therefore a tradeoff between maximizing site inclusion (higher β_{sim}) and reducing variance (lower β_{sim}). As the goal for our first analysis (Level 4a) was to assess general patterns of species similarity, we cut the dendrogram at a β_{sim} value of 0.65. This resulted in three major clusters that contained the vast majority (3596/3615, 99.5%) of the sites and produced spatially-coherent clusters (see results), while keeping the variance in cluster size low (Figure 5). To examine finer scale ecological units, we cut the tree at a β_{sim} value of 0.55 to obtain Level 4b clusters. This distance cut-off resulted in five clusters with less variance among clusters (more evenness in cluster size; (Figure 4, Figure 5), but with slightly fewer sites represented overall in the top 5 clusters (3499/3615, 96.8%) than in the Level 4a cutoff. All cluster analyses were carried out using R packages 'vegan', 'simba', 'maptools', 'dendroextras', and supporting packages (R Core Development Team 2013).

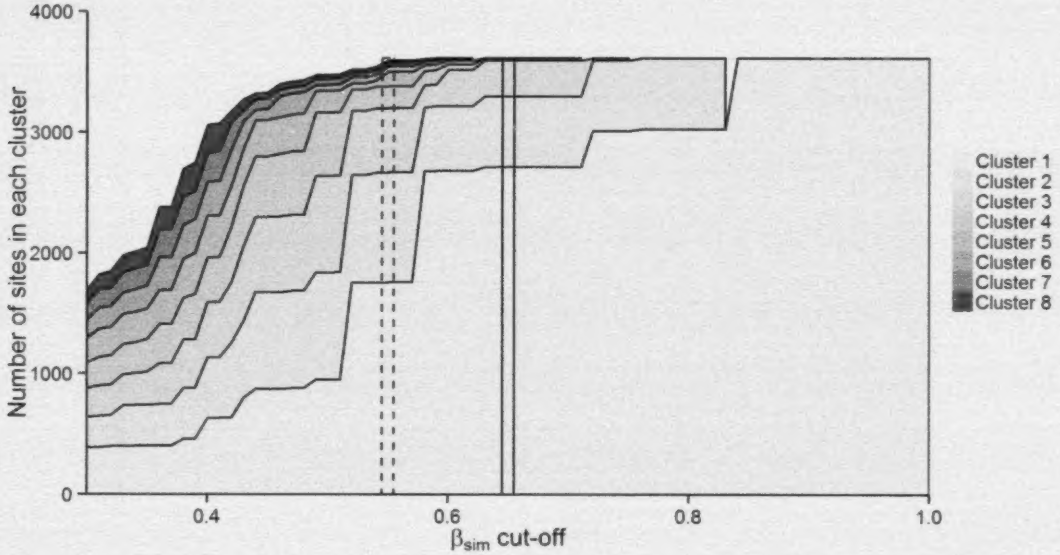


Figure 4. Number of sites (out of a possible 3615) in each of the top 8 clusters for dendrograms cut at increasing β_{sim} values. The dashed-line and solid boxes show cluster distribution at β_{sim} cutoffs of 0.55 and 0.65, respectively. Dendrograms cut with a β_{sim} of 1 contain only one cluster, which contains all sites. Lower β_{sim} values result in more clusters, but fewer sites retained in the most populated clusters.

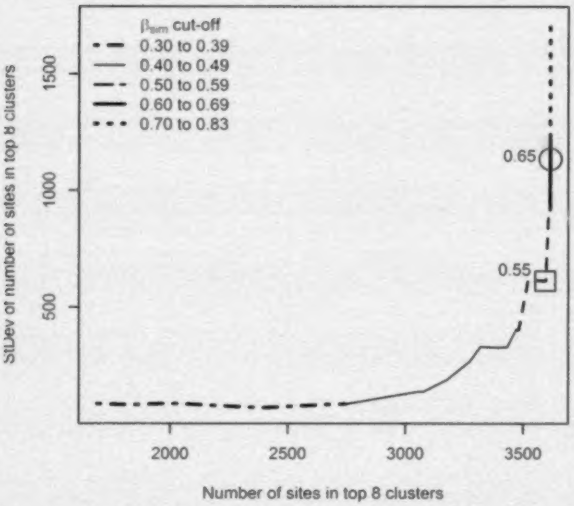


Figure 5. Variance (standard deviation) of the number of sites in the top 8 clusters compared to the number of sites retained in those clusters for increasing β_{sim} cut-off values. β_{sim} values of 0.55 (square) and 0.65 (circle) show variance and number of sites for the biophysical unit analyses, respectively. Higher β_{sim} cut-off values increase the number of sites included in the top clusters (see Figure 4), but reduces the evenness of those clusters.

3.2.3 Indicator species analysis

In order to identify species-habitat associations, we conducted an indicator species analysis (Dufrêne and Legendre 1997) to determine which species were most prevalent in each cluster at the two biophysical unit levels (Level 4a and b). The indicator species analysis was run on the clusters using the R function *Indval* in the "LabDSV" package. *Indval* calculates an indicator value for each species, ranging from 0 to 1, based on the relative frequency of each species in a cluster compared to those values for that species in all other clusters, and therefore can be interpreted as how prevalent that species is within the cluster. A permutation test also calculates a p-value for each indicator value and species. Although *Indval* was designed for abundance data, it performs well for presence-absence data (Podani and Csányi 2010). We report indicator species within each cluster that were significant ($p < 0.05$) and had an *Indval* value of > 0.25 (following Dufrêne and Legendre 1997). The term 'indicator species' in this analysis means a species that is most prevalent in a particular cluster in comparison to other clusters. It does not identify species that are good indicators of ecosystem health as the term is often used.

3.2.4 Environmental correlates of biological clusters

A random forest analysis was carried out to determine if environmental variables could explain the variation in the biological clusters, and whether Level 4a and b could be accurately predicted in areas with no biological data. Random forest is a machine-learning method that creates an ensemble or "forest" of classification trees. It avoids developing a tree model that is over-fit to the training data by using bootstrap aggregation or "bagging" to repeatedly sample the data with replacement (bootstrapping) and developing trees for each dataset (Cutler et al. 2007). The "out of the bag" sample (OOB, about 1/3 of the data) are held out of the resampling and used to evaluate the model (Franklin 2009) using a metric analogous to R^2, called pseudo R^2. Here we use cluster membership as the response variable to determine if environmental variables can be used to accurately predict clusters, and if so, use these relationships to delineate the Level 4 boundaries and predict which ecological unit each site belongs to in areas with no biological data. Key advantages of the random forest method are that is does not overfit models, it can be used to model non-linear and threshold responses, and it ranks the importance of variables for classification prediction (Franklin 2009).

Environmental rasters were resampled from their original spatial resolution (Table 5) to a 4 km resolution to match the biological data. Although random forest can handle correlated variables, we conducted a preliminary correlation analysis on the 59 environmental variables to select a reduced set in the random forest model. Many of the environmental variables are multiple measurements of the same variable, for example, there are eight bottom temperature variables including fall, summer, winter, spring, mean, max, min, and range (see Table 5). To aid in interpretability, we chose one or two representatives for each environmental category that were representative of that variable and were not highly correlated with other variables in our analysis. We also chose variables that had coverage for the entire study area, and were most likely to be biologically relevant. We retained 14 variables that were not highly correlated (R^2 for each pair of variables < 0.7, Appendix C) for the random forest analysis, including depth, rugosity, slope, flow (summer and winter), tidal direction, tidal speed, bottom salinity (summer), bottom temperature range, sea surface temperature (overall), and surface values for nitrate, dissolved oxygen, phosphate, and silicate. It is important to note that the sampling period for our biological data (~2003-2013), matched when possible, the sampling period for the environmental data and that this sampling period occurred during a negative phase of the Pacific Decadal Oscillation (NOAA PDO). Although the PDO is more likely to impact pelagic patterns, the outputs of our benthic analysis will reflect only one phase of the PDO.

We examined the number of surveys at each site to assess the effect of uneven sampling over our study area on cluster membership. Because of the different sampling methods within each data source, these "surveys" are not standardized but represent a semi-quantitative measure of sampling effort across the region. Depending on the data source, a survey could be a trap or trap line, a trawl set, or a longline set. A map of sampling effort (Figure 6) shows that sampling effort was relatively uniform with patchy areas of higher effort. Sampling effort does not appear strongly correlated with species richness at each site (Figure 7); however, to assess the role of survey effort in driving the patterns observed in the cluster analysis, we also included number of surveys per site as a predictor in the random forest analysis.

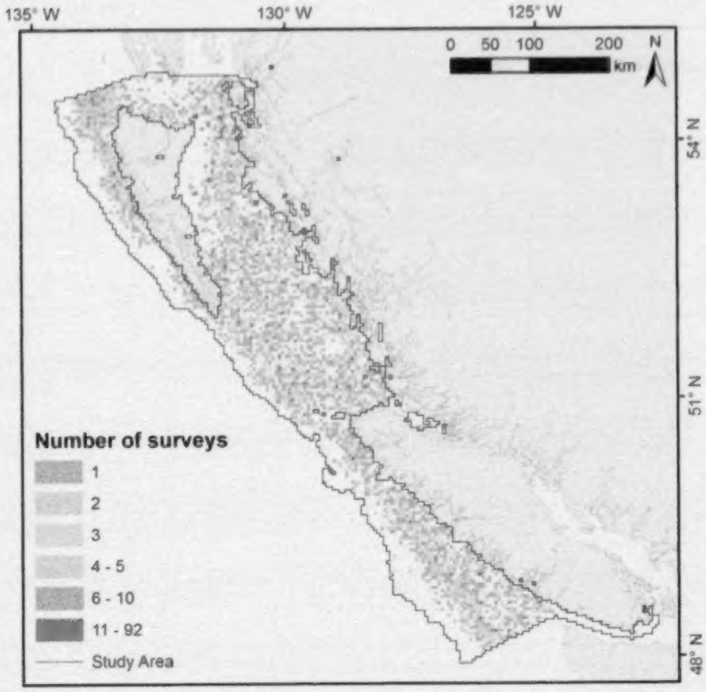

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Number of surveys
1
2
3
4 - 5
6 - 10
11 - 92
Study Area

Figure 6. Sampling effort for each sampling site (4 km grid cell) for data included in the cluster analysis and random forest analysis. The definition of an individual survey depended on the dataset (Tanner Crab, Crab Trap, or Groundfish),and included trawl tows, longline sets, traps, and trap sets.

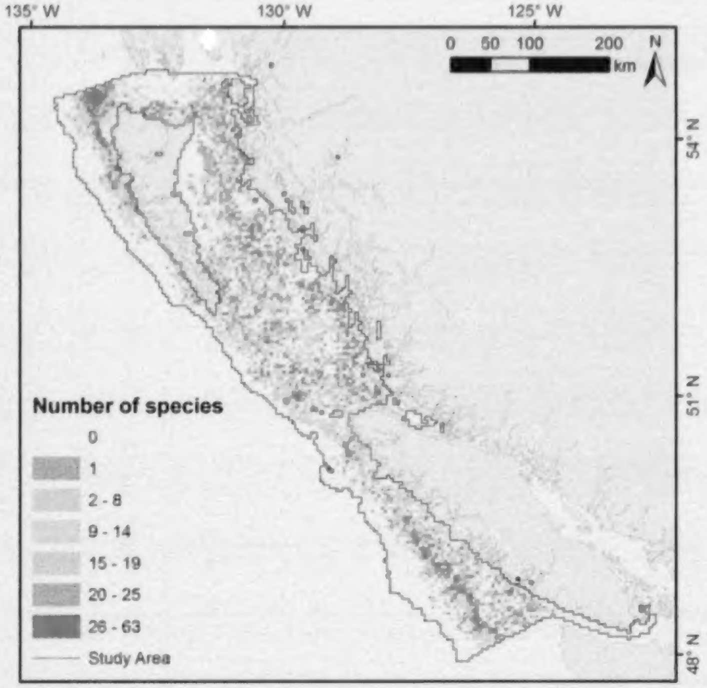


Figure 7. Number of species recorded at each sampling site (4 km grid cell) for the dataset used in the cluster analyses. Sites with only one species recorded were not included in the cluster analyses (see Section 3.2.2 for details).

Of the 3615 sites input into the cluster analysis, 3596 sites were assigned to a major Level 4a cluster, and of those sites, 3593 had values for the above 15 predictor variables (14 environmental layers plus number of surveys); for Level 4b, 3499 sites were assigned to a major cluster, 3496 of which had values for all 15 predictors. To assess the role of sampling effort on cluster membership, a random forest model (10,000 trees) was run on the full dataset (3593 and 3496 sites for Levels 4a and 4b, respectively) using all 15 variables as predictors. The results of this analysis indicated that survey effort had little to no effect on cluster membership: the mean decrease in accuracy for survey effort, overall and for each biophysical unit, was < 0.2%, the lowest of any predictor variable. Therefore the number of surveys per site was dropped from further analyses.

Model parameters and performance metrics

The random forest model was implemented in R, using the randomForest package (Liaw and Wiener 2002, R Core Development Team 2013) and 10,000 trees were generated for each run. The accuracy of the model was assessed as 100 – [% out-of-bag error], and model performance was also assessed using 10-fold cross validation, where the input data were randomly divided into ten subsamples, and each subsample (10% of full dataset) was used to test the prediction accuracy of a 10,000-tree model built on the remaining data (90%). Model fits were quantified using the area under the receiver operating characteristic curve (AUC). AUC values typically range from 0.5 for classifiers that perform no better than random to 1.0 for perfect classification (Fawcett 2006). AUC values from each cross-validation run ($n = 10$) were averaged to assess overall fit of the model. The relative importance of each predictor variable was also obtained from the cross-validation analysis, by taking the average of the mean decrease in model accuracy for each predictor for each 90-10 split. The variable importance plots were examined to assess the importance of each predictor in the classification of each biological cluster individually, and for the overall model.

The results of the random forest model were projected onto the 14 predictor environmental rasters to get a surface of predicted cluster membership for sites that were not used to create the model (i.e., sites not assigned to a major cluster and sites without biological data). Of sites within our study area that contained all 14 environmental variables (6856 of 6875 sites), 47% (3241) had no biological data but we used the model to predict which cluster was present at each site. Using the relationships between the environmental data and the biological assemblage data, we delineated ecological units to Level 4a and 4b, respectively. The ecological unit refers to the biological assemblage present and predicted by our model, for the entire study area. In addition to the internal model accuracy assessment (out-of-bag error), and the 10-fold cross-validation, we examined uncertainty in the model by mapping the percentage of votes underlying each predicted cluster classification (the output of the random forest model). This evaluation provides a visualization of the underlying uncertainty in the predicted surface and identifies areas in the classification with higher/lower relative confidence.

As a further test of the choice of dendrogram distance cut-off, we ran a sensitivity analysis on fit (AUC) for models developed from trees cut at all distances from β_{sim} = 0.35 to 0.90. For each cutoff, we identified the "major clusters" as those that cumulatively included 90% of the sites in the analysis (> 3253 sites); these clusters were used as the response variable (classes), with the above 14 environmental variables as predictors. The number of clusters tested ranged from two for β_{sim} values between 0.63 and 0.83, to 29 for β_{sim} = 0.35 (Figure 8). For the outputs of each β_{sim} cut-off, 10-fold cross validation (see above for description) was carried out and average AUC was calculated. Model fit was highest when fewer clusters were tested (i.e., at higher β_{sim} values, Figure 8), with decreasing fit at lower β_{sim} values. The mean AUC was greater than 0.98 (excellent fit) for β_{sim} values > 0.57, with very good fit (> 0.90) for β_{sim} values between 0.44 and 0.56. These results support our decision to use β_{sim} cut-offs of 0.65 and 0.55 for our Level 4a and b, respectively.

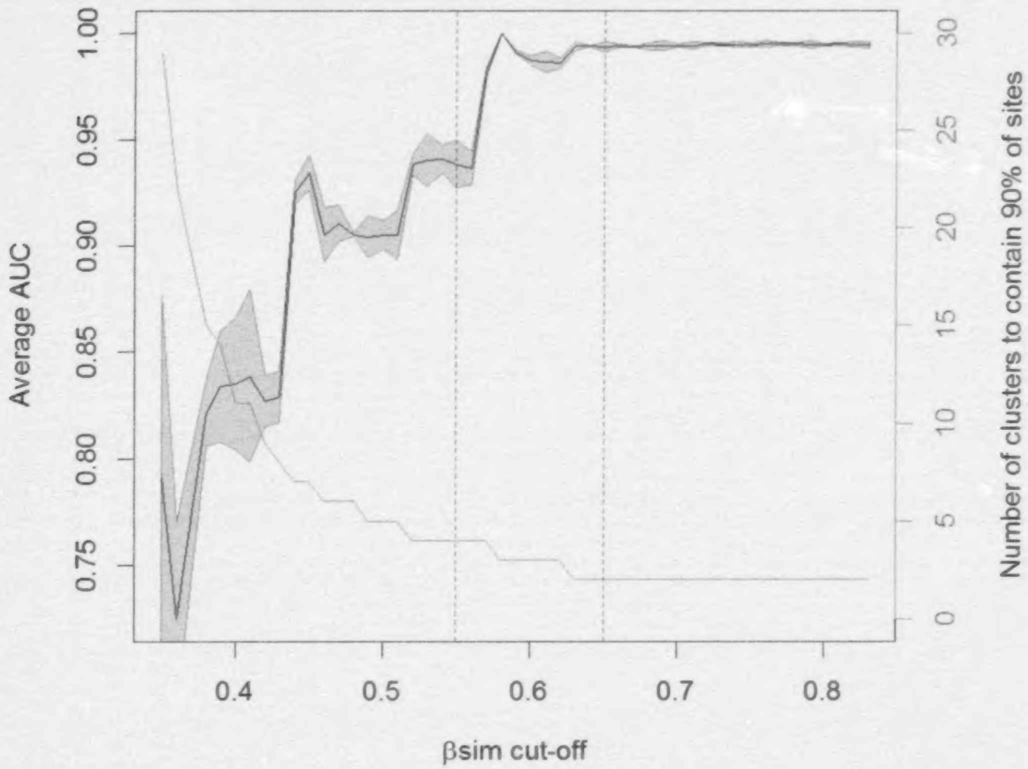


Figure 8. Model fit (AUC) for random forest models predicting membership to clusters resulting from dendrograms cut at different β_{sim} values. Grey area indicates standard deviation around the mean for 10 model runs (with random 90-10% training and testing datasets) at each β_{sim} value. Red line shows the number of clusters (classes) tested at each β_{sim} value. Blue lines indicate $\beta_{sim} = 0.55$ and $\beta_{sim} = 0.65$ (biophysical units analyses).

3.3 CLUSTER ANALYSIS RESULTS

A cluster analysis was run on the 174 species from 3615 sites (See Appendix D for list of species included and their frequencies).

3.3.1 Description of clusters

3.3.1.1 Level 4a: Large-scale Biophysical units

Level 4a clusters ($\beta_{sim} = 0.65$) represent areas in the Pacific Region with the largest differences in species composition. The dendrogram cut at this level classifies all of the sites into five clusters, with 99.5% of the sites assigned to the three most populous clusters (Figure 10): the Shelf (2710 sites), Slope (588 sites), and Banks (298 sites; Figure 9). The two clusters with very low membership (1 and 18 sites, totalling 0.5% of sites; "unclassified" sites on Figure 9) were not considered further. The cluster with one site was found on Goose Island Bank, surrounded by "Bank"-classified sites. Of the 18 sites in the other cluster, two were found in Cook Trough and the rest were scattered off the west coast of Vancouver Island, all adjoining "Shelf"-classified sites.

The three major clusters cover spatially coherent areas:

1. the Shelf cluster covers most of the continental shelf including Queen Charlotte Sound, Dixon Entrance, and most of the shelf areas off the west coast of Vancouver Island;
2. the Slope cluster includes sites on the continental slope, stretching northwest to southeast; and
3. the Banks cluster includes several shallower shelf regions, including Cook Bank, Goose Bank, Dogfish Bank, and some sites to the west of Vancouver Island.

For the three Level 4a clusters, indicator species analysis *Indval* returned 157 species that were significant at $p < 0.05$ (permutational test in *Indval)*. An *Indval* value cut-off of 0.25 left 42 species associated with a large-scale biophysical cluster as reported in Table 6. A high *Indval* value indicates that a species is not only very frequent in that particular group, but also that it is infrequent in other clusters indicating the level of group fidelity. The Slope cluster has the highest *indval* values of any cluster with three species with *Indval* values of over 0.7 (Giant Grenadier, *Albatrossia pectoralis*; Grooved Tanner Crab, *Chionoecetes tanneri;* and Pacific Grenadier, *Coryphaenoides acrolepis)* suggesting that these three species have a strong association with Slope habitat. The Banks cluster has the second highest *indval* values with the highest being 0.699 for Rock Sole (*Lepidopsetta bilineata*), whereas the Shelf cluster's highest *Indval* value was for Arrowtooth Flounder, *Atheresthes stomias* at 0.565. Arrowtooth Flounder had a high frequency in the Shelf cluster (85.3%), but its *Indval* value of 0.565 indicates that it is also present in other clusters, albeit in much lower frequencies. In contrast, Giant Grenadier has a high frequency in the Slope cluster (74.7%) and also a very high *Indval* value (0.734) indicating that it is rarely found in other clusters.

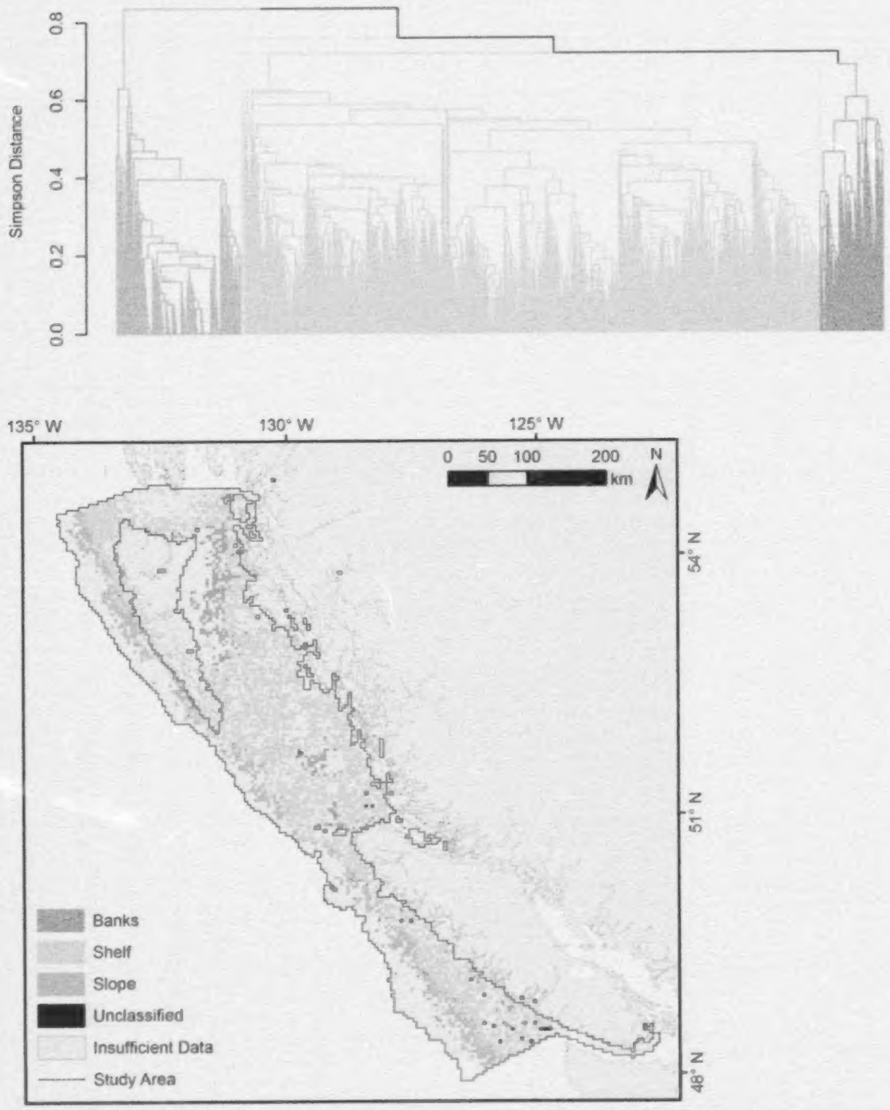

Simpson Distance
0.0
0.2
0.4
0.6
0.8
135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Banks
Shelf
Slope
Unclassified
Insufficient Data
Study Area

Figure 9. Dendrogram (top) cut at β_{sim} height of 0.65 and resulting map (bottom) of sites belonging to each cluster. "Unclassified" sites are shown in grey on the dendrogram and black squares on the map.

Table 6. Indicator species for Level 4a (large-scale Biophysical units) clusters produced by cluster analysis of species assemblages of demersal fish and invertebrates listed in order of Indval metric. All Indval values reported here are significant (p<0.05) using a permutation test. Taxonomic names shown are those used throughout our analyses, as cross-referenced with the World Register of Marine Species.

Top cluster	Species name	Common Name	In top cluster		Frequency in other clusters		
			Indval	Freq	Shelf	Slope	Banks
Shelf	*Atheresthes stomias*	Flounder, Arrowtooth	0.565	85.3%	-	19.2%	24.2%
	Glyptocephalus zachirus	Sole, Rex	0.476	64.9%	-	8.5%	15.1%
	Sebastes brevispinis	Rockfish, Silvergray	0.454	51.9%	-	2.0%	5.4%
	Raja rhina	Skate, Longnose	0.413	57.2%	-	8.7%	13.4%
	Sebastes babcocki	Rockfish, Redbanded	0.412	46.8%	-	5.3%	1.0%
	Sebastes alutus	Perch, Pacific Ocean	0.389	44.7%	-	6.0%	0.7%
	Microstomus pacificus	Sole, Dover	0.344	59.6%	-	27.6%	16.1%
	Lyopsetta exilis	Sole, Slender	0.334	39.0%	-	2.6%	4.0%
	Strongylocentrotus fragilis	Sea Urchin, Pink	0.291	37.1%	-	8.8%	1.3%
	Gadus macrocephalus	Cod, Pacific	0.289	57.2%	-	0.5%	55.4%
	Ophiodon elongatus	Lingcod	0.286	54.7%	-	2.2%	47.7%
	Sebastes elongatus	Rockfish, Greenstriped	0.283	30.6%	-	1.2%	1.3%
	Eopsetta jordani	Sole, Petrale	0.280	40.7%	-	0.0%	18.5%
	Theragra chalcogramma	Pollock, Walleye	0.265	41.7%	-	0.2%	23.8%
	Sebastes ruberrimus	Rockfish, Yelloweye	0.261	33.3%	-	0.9%	8.4%
	Sebastes zacentrus	Rockfish, Sharpchin	0.255	28.7%	-	1.9%	1.7%
	Pandalus jordani	Shrimp, Pink	0.253	27.9%	-	0.2%	2.7%
Slope	*Albatrossia pectoralis*	Grenadier, Giant	0.734	74.7%	1.3%	-	0.0%
	Chionoecetes tanneri	Crab, Grooved Tanner	0.731	74.7%	1.6%	-	0.0%
	Coryphaenoides acrolepis	Grenadier, Pacific	0.726	73.6%	1.1%	-	0.0%
	Anoplopoma fimbria	Sablefish	0.526	91.3%	58.7%	-	8.4%
	Sebastolobus altivelis	Thornyhead, Longspine	0.432	45.6%	2.5%	-	0.0%
	Sebastolobus alascanus	Thornyhead, Shortspine	0.398	61.4%	33.3%	-	0.0%
	Lithodes couesi	Crab, Scarlet King	0.397	40.8%	1.2%	-	0.0%
	Coryphaenoides cinereus	Grenadier, Popeye	0.257	25.9%	0.2%	-	0.0%
Banks	*Lepidopsetta bilineata*	Sole, Rock	0.699	83.2%	15.8%	0.0%	83.2%
	Psettichthys melanostictus	Sole, Pacific Sand	0.502	51.7%	1.5%	0.0%	51.7%
	Hydrolagus colliei	Ratfish, Spotted	0.448	85.9%	76.4%	2.4%	85.9%
	Ammodytes hexapterus	Sand lance, Pacific	0.426	46.3%	4.0%	0.0%	46.3%
	Raja binoculata	Skate, Big	0.386	51.3%	16.9%	0.0%	51.3%
	Hippoglossus stenolepis	Halibut, Pacific	0.383	75.2%	66.9%	5.4%	75.2%
	Pleuronichthys decurrens	Sole, Curlfin	0.374	42.6%	6.0%	0.0%	42.6%
	Podothecus accipenserinus	Poacher, Sturgeon	0.366	40.3%	4.1%	0.0%	40.3%
	Parophrys vetulus	Sole, English	0.355	57.4%	34.6%	0.7%	57.4%

Top cluster	Species name	Common Name	In top cluster		Frequency in other clusters		
			Indval	Freq	Shelf	Slope	Banks
	Pisaster brevispinus	Sea Star	0.349	36.9%	2.2%	0.0%	36.9%
	Chitonotus pugetensis	Sculpin, Roughback	0.347	35.9%	1.3%	0.0%	35.9%
	Sebastes maliger	Rockfish, Quillback	0.312	45.0%	19.9%	0.0%	45.0%
	Pycnopodia helianthoides	Sea Star	0.311	40.3%	11.5%	0.3%	40.3%
	Hexagrammos decagrammus	Greenling, Kelp	0.301	32.6%	2.7%	0.0%	32.6%
	Citharichthys sordidus	Sanddab, Pacific	0.298	41.3%	15.6%	0.2%	41.3%
	Metacarcinus magister	Crab, Dungeness	0.287	30.2%	1.6%	0.0%	30.2%
	Sebastes caurinus	Rockfish, Copper	0.259	27.9%	2.1%	0.0%	27.9%

3.3.1.2 Level 4b: Biophysical units

Cutting the tree slightly lower, at $\beta_{sim} = 0.55$, retained a high proportion of sites assigned to the most populous clusters (96.8% of sites in the top 5 clusters; 99.4% in the top 8) and revealed smaller areas of spatially coherent species assemblages that we call Level 4b – Biophysical units (Figure 10).

Two more distinct assemblages became apparent in the Level 4b compared to Level 4a: the Banks unit is split to distinguish Dogfish Bank from Other Banks, while the Shelf unit is split into the Troughs and Shelf units. The Slope unit is very similar to the Slope large-scale biophysical unit, with slightly fewer sites being captured at the lower cutoff.

For the clusters resulting from cutting the tree at $\beta_{sim} = 0.55$, the indicator species analysis *Indval* returned 163 species that were significant at $p < 0.05$ (permutational test in *Indval)*. An *Indval* value cut-off of 0.25 left 34 species associated with a Level 4b cluster (Table 7). Twelve species were significantly associated with the Dogfish Banks cluster, 10 with the Troughs cluster, six with the Slope cluster, four with the Other Banks cluster, and two with the Shelf cluster (Table 7). The Slope cluster at Level 4b is very similar to the slope cluster at Level 4a, and had the highest *Indval* values of any cluster with the same three species with high *Indval* values (*Indval* = 0.761 for Grooved Tanner Crab, *Chionoecetes tanneri; Indval* = 0.700 for Giant Grenadier, *Albatrossia pectoralis;* and *Indval* = 0.695 for Pacific Grenadier, *Coryphaenoides acrolepis)*. The Dogfish bank cluster had the second highest *Indval* values of any Level 4b cluster with the highest being 0.716 for the Pacific Sand Sole (*Psettichthys melanostictus*), whereas the Troughs cluster's highest *Indval* value was 0.54 for both the Redbanded Rockfish (*Sebastes babcocki*) and Pacific Ocean Perch (*Sebastes alutus*). The Yelloweye Rockfish (*Sebastes ruberrimus*) had the highest frequency in the Shelf cluster (45.92%), but its *Indval* value of 0.278 indicates that it is also present in other clusters, although in lower frequencies.

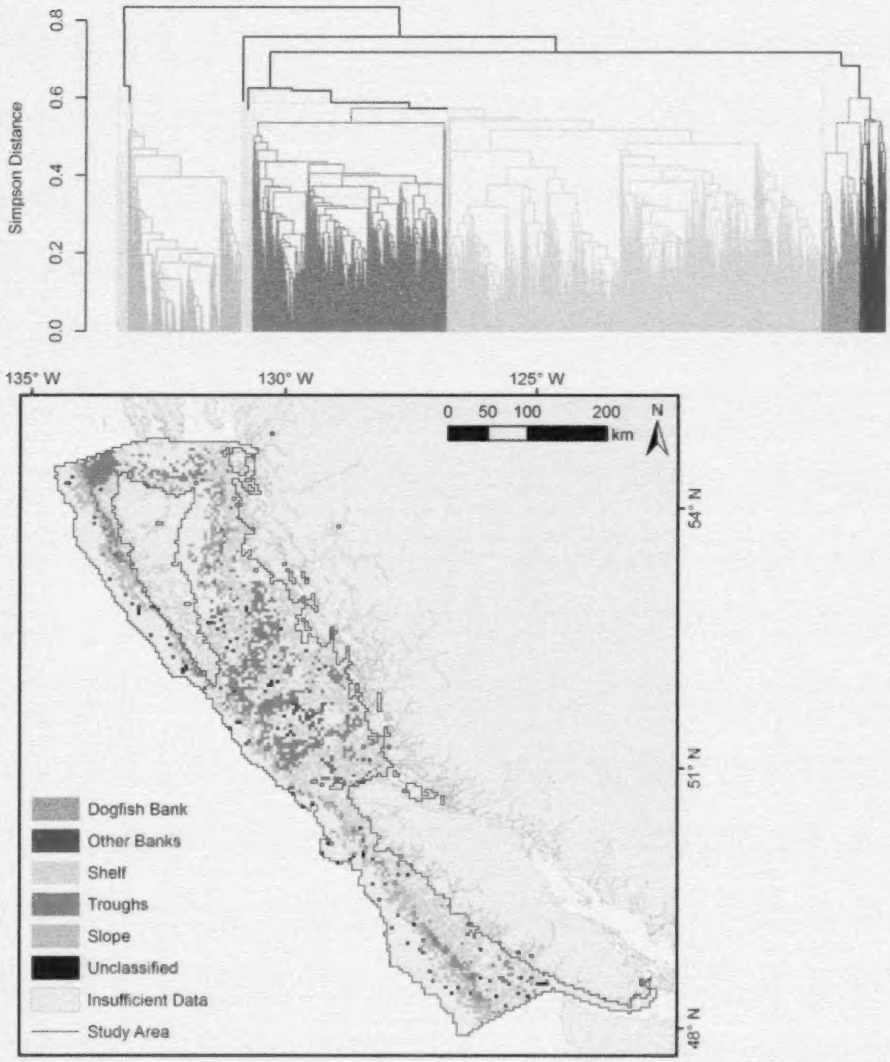

Simpson Distance
0.0
0.2
0.4
0.6
0.8
135° W
130° W
125° W
0
50
100
200
km
N
54° N
51° N
48° N
Dogfish Bank
Other Banks
Shelf
Troughs
Slope
Unclassified
Insufficient Data
Study Area

Figure 10. Dendrogram (top) cut at β_{sim} height of 0.55 and resulting map (bottom) of sites belonging to each cluster. "Unclassified" sites are shown in grey on dendrogram and black squares on map.

Table 7. Indicator species for Level 4b clusters produced by cluster analysis of species assemblages of demersal fish and invertebrates listed in order of Indval metric. All Indval values reported here are significant ($p<0.05$) using a permutation test.

Top clusters	Species name	Common name	In top cluster		Frequency in other clusters				
			Indval	Freq.	Shelf	Troughs	Slope	Dogfish Bank	Other Banks
Shelf	*Sebastes ruberrimus*	Rockfish, Yelloweye	0.278	45.9%	-	8.1%	0.9%	0.6%	20.3%
	Eopsetta jordani	Sole, Petrale	0.255	52.9%	-	17.6%	0.0%	13.9%	25.4%
Troughs	*Sebastes alutus*	Perch, Pacific Ocean	0.541	78.7%	27.5%	-	6.6%	0.0%	1.7%
	Sebastes babcocki	Rockfish, Redbanded	0.540	80.0%	30.3%	-	5.8%	0.6%	1.7%
	Sebastes aleutianus	Rockfish, Rougheye	0.496	68.2%	8.4%	-	17.1%	0.0%	0.0%
	Pandalopsis dispar	Shrimp, Sidestripe	0.398	49.0%	8.5%	-	1.7%	1.1%	0.0%
	Sebastolobus alascanus	Thornyhead, Shortspine	0.389	77.6%	10.9%	-	66.0%	0.0%	0.0%
	Stheresthes stomias	Flounder, Arrowtooth	0.379	96.3%	80.2%	-	21.2%	26.7%	20.3%
	Microstomus pacificus	Sole, Dover	0.366	83.6%	48.1%	-	29.3%	21.1%	8.5%
	Glyptocephalus zachirus	Sole, Rex	0.360	79.8%	57.9%	-	9.4%	16.1%	13.6%
	Strongylocentrotus fragilis	Sea Urchin, Pink	0.341	57.8%	26.9%	-	9.8%	0.0%	3.4%
	Sebastes diploproa	Rockfish, Splitnose	0.271	32.9%	4.2%	-	2.8%	0.0%	0.0%
Slope	*Chionoecetes tanneri*	Crab, Grooved Tanner	0.761	80.3%	0.4%	3.9%	-	0.0%	0.0%
	Albatrossia pectoralis	Grenadier, Giant	0.700	73.7%	0.0%	3.8%	-	0.0%	0.0%
	Coryphaenoides acrolepis	Grenadier, Pacific	0.695	72.2%	0.4%	2.4%	-	0.0%	0.0%
	Lithodes couesi	Crab, Scarlet King	0.412	44.4%	0.1%	3.3%	-	0.0%	0.0%
	Sebastolobus altivelis	Thornyhead, Longspine	0.407	46.8%	0.4%	6.6%	-	0.0%	0.0%
	Anoplopoma fimbria	Sablefish	0.392	97.9%	45.5%	86.0%	-	11.7%	3.4%
Dogfish Bank	*Psettichthys melanostictus*	Sole, Pacific Sand	0.716	80.6%	2.4%	0.0%	0.0%	-	7.6%
	Podothecus accipenserinus	Poacher, Sturgeon	0.502	61.7%	5.9%	0.5%	0.0%	-	7.6%
	Metacarcinus magister	Crab, Dungeness	0.449	48.9%	2.2%	0.4%	0.0%	-	1.7%
	Pisaster brevispinus	Sea Star	0.405	52.8%	3.4%	0.0%	0.0%	-	12.7%
	Lepidopsetta bilineata	Sole, Rock	0.386	85.6%	22.9%	1.5%	0.0%	-	79.7%
	Parophrys vetulus	Sole, English	0.386	79.4%	45.5%	14.2%	0.8%	-	23.7%
	Pycnopodia helianthoides	Sea Star	0.370	57.8%	17.0%	1.4%	0.4%	-	13.6%
	Isopsetta isolepis	Sole, Butter	0.361	43.3%	5.1%	0.1%	0.0%	-	3.4%
	Chitonotus pugetensis	Sculpin, Roughback	0.347	48.3%	1.9%	0.1%	0.0%	-	16.9%
	Microgadus proximus	Tomcod, Pacific	0.327	44.4%	8.9%	0.3%	0.0%	-	6.8%
	Raja binoculata	Skate, Big	0.318	62.8%	24.6%	2.7%	0.0%	-	33.9%
	Lumpenus sagitta	Prickleback, Snake	0.318	37.8%	4.2%	0.4%	0.0%	-	2.5%
Other Banks	*Sebastes maliger*	Rockfish, Quillback	0.427	74.6%	29.5%	0.8%	0.0%	25.6%	-
	Hexagrammos decagrammus	Greenling, Kelp	0.322	49.2%	4.1%	0.1%	0.0%	21.7%	-
	Chlamys rubida	Scallop, Pink	0.281	37.3%	4.7%	0.3%	0.0%	7.2%	-
	Hydrolagus colliei	Ratfish, Spotted	0.254	90.7%	82.5%	65.1%	2.6%	82.8%	-

3.3.2 Environmental differences among biophysical units

3.3.2.1 Level 4a: Large-scale Biophysical units

The predictor variables in our model accurately differentiated the three major large-scale biophysical units. The random forest analysis showed high accuracy with an out-of bag (OOB) misclassification rate of only 7.15 % (pseudo $R^2 = 92.85$), The predictive power of the model was high, with the AUC for cross-validation of the random forest model at 0.989 ± 0.005. An AUC value greater than 0.9 indicates high model performance, in that clusters are well-explained by the environmental variables included in the model.

The importance values (Figure 11) show that depth, salinity, and temperature range are the most important environmental parameters overall for distinguishing Level 4a clusters. Depth and salinity were the most important variables for the overall model, however examining each Level 4a unit separately showed differences in relative variable importance. The Banks large-scale Biophysical unit is more highly related to salinity, whereas depth was more important for the Slope large-scale Biophysical unit. Although not ranked very high in the overall model, tidal speed was an important predictor for both the Banks and Slope large-scale Biophysical units, particularly for the Slope. Banks are shallower, have a larger temperature range, and have lower salinity than the Shelf and Slope large-scale Biophysical units, whereas the Slope is much deeper, has a more stable temperature, and has higher salinity (Figure 12).

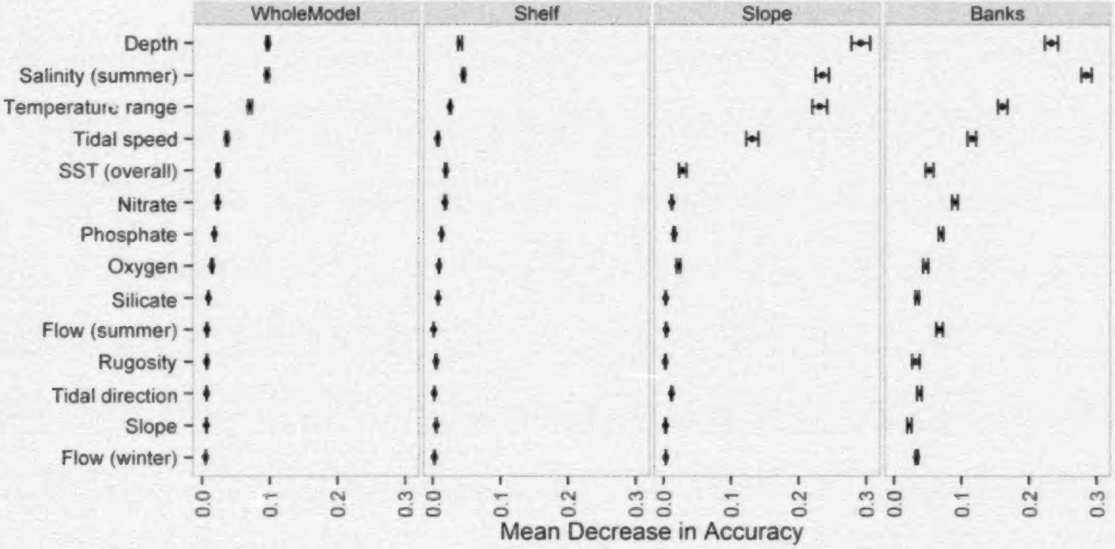


Figure 11. Variable importance (mean ± SD decrease in accuracy) for random forest model (ten runs of random 70% of data, 10,000 trees each run) for whole model and for each large-scale biophysical unit individually.

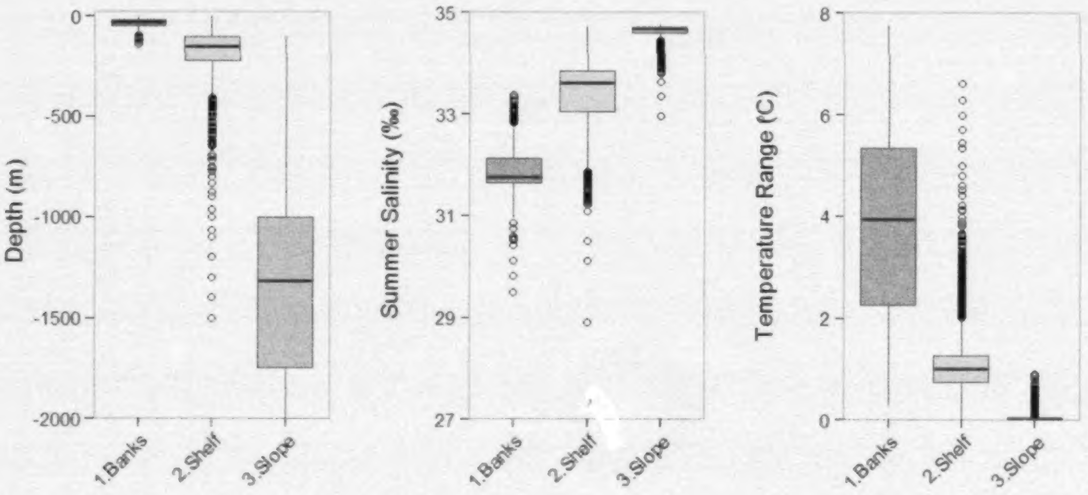


Figure 12. Boxplots of most important environmental variables for differentiating the three large-scale biophysical units, as determined by a random forest analysis.

The random forest model was used to predict the Level 4a unit present in areas with no biological information in our study area (3241 of the 4 km grid cells, or 47% of study area) using information based on the environmental conditions at those sites (Figure 13). Although the overall model AUC was very high (0.989) indicating excellent model performance, it is also important to understand areas where the model is not performing as well. We mapped uncertainty (quantified by the percentage of votes to designated cluster, the measure used in the random forest model) to indicate areas with higher uncertainty (Figure 14). The uncertainty map shows that areas associated with boundaries of each biophysical unit often have a lower percentage of votes in the random forest model than areas within the core of each biophysical unit. We interpret this finding to mean that the model does not perform as well in transition zones, where species composition is changing across an environmental gradient.

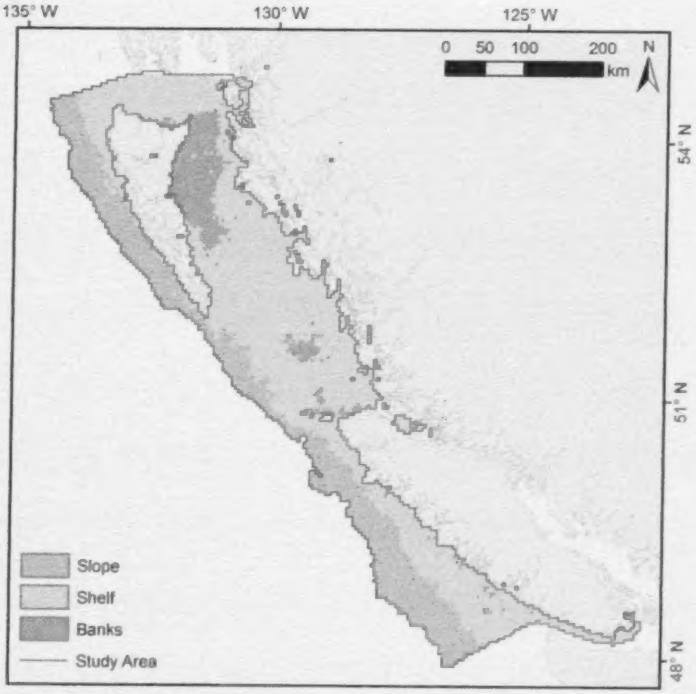

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Slope
Shelf
Banks
Study Area

Figure 13. Random forest model-predicted output for site classification into major clusters (Level 4a, large-scale Biophysical units).

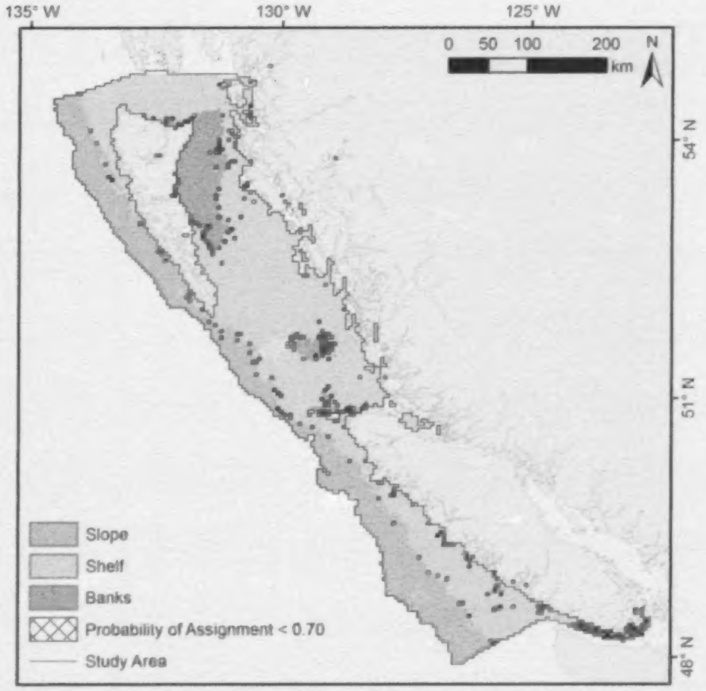


Figure 14. Level 4a (large-scale Biophysical units) classification showing sites with low (< 0.70) probability of assignment from the random forest analysis. Probabilities are calculated as the proportion of trees in which a grid cell is assigned to a class ("votes") out of the number of trees assembled in the random forest (10,000).

3.3.2.2 Level 4b: Biophysical units

The environmental variables included in the random forest model accurately classified each cluster with an out-of bag misclassification rate of 15.96% (pseudo R^2 = 84.04). The predictive power of the model was high, with the AUC for cross-validation of the random forest model at 0.93 ± 0.01. The variable importance plots (Figure 15) were similar to the Level 4a plots and show that depth, salinity, and temperature range were also the most important environmental parameters among those evaluated for differentiating the clusters. When examining each cluster separately, temperature range appears to be less important for the Other Banks and Trough Clusters, than it is for all other clusters. Tidal speed is an important predictor for Other Banks and the Slope clusters and nitrate is a more important predictor for the Dogfish Bank cluster than for the other clusters. Boxplots of the top three environmental predictors show that Dogfish Bank is the shallowest biophysical unit and experiences the greatest range in temperature (Figure 16).

Using the relationships between the environmental data and the biological assemblage data, Biophysical units at Level 4b were predicted across the study area (Figure 17). Although the overall model AUC was very high (0.93), we mapped uncertainty (quantified by the percentage of votes to designated cluster) to highlight the underlying uncertainty in the model (Figure 18). This uncertainty map is similar to the Level 4a uncertainty analysis map, showing that the areas associated with boundaries between biophysical units had a lower percentage of votes in the random forest model than areas within the core of each unit. This finding is particularly true at the southern boundary of Dogfish Bank, around the Other Banks, and running along the length of the shelf break between the Shelf and Slope. Based on these results, we suggest that the model does not perform as well in transition zones, where the species composition is changing across an environmental gradient. An additional area with high uncertainty in our model is around the southern end of the Southern Shelf Bioregion in the Juan de Fuca Strait around the tip of Vancouver Island. Due to the influences of local currents and eddies in the area, it is likely that our broad-scale abiotic data do not accurately capture the environmental complexity of this area leading to poor model performance in this area. This result supports our decision to remove sites in close proximity to land, and the Strait of Georgia Bioregion, and provides evidence that these areas should be modeled separately and at a local scale with regionally collected biotic and abiotic data if available.

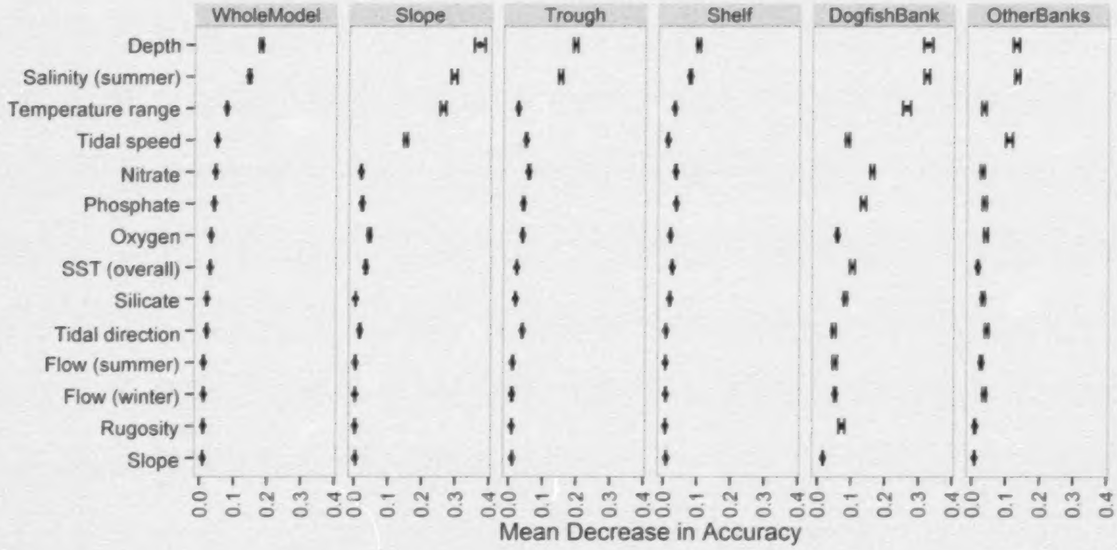


Figure 15. Variable importance (mean ± SD decrease in accuracy) for random forest model (ten runs of random 70% of data, 10,000 trees each run) for whole model and for each biophysical unit individually.

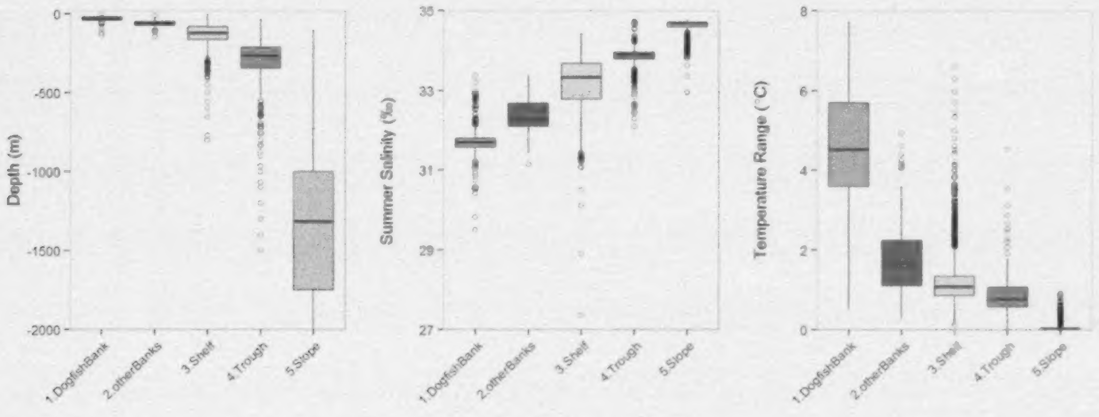


Figure 16. Boxplots of most important environmental variables for differentiating the five biophysical units, as determined by a random forest analysis.

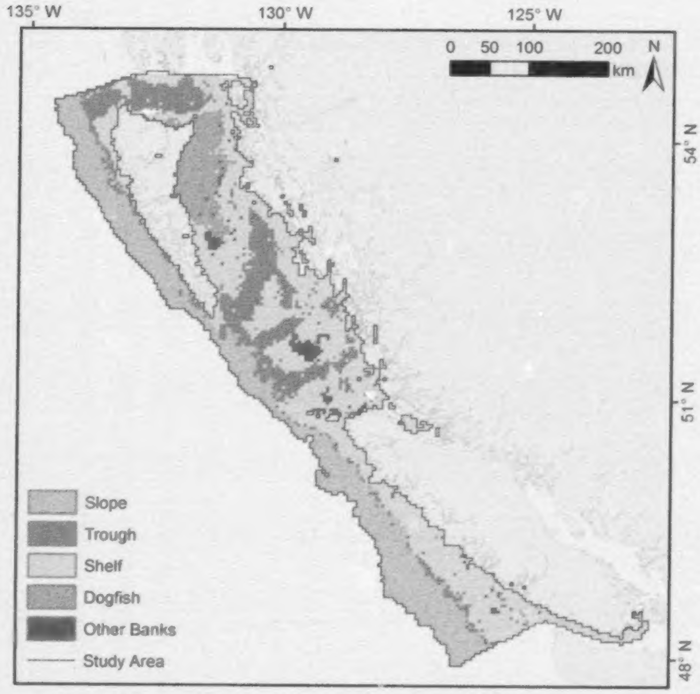

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Slope
Trough
Shelf
Dogfish
Other Banks
Study Area

Figure 17. Random forest model-predicted output for site classification into major clusters (Level 4b, biophysical units).

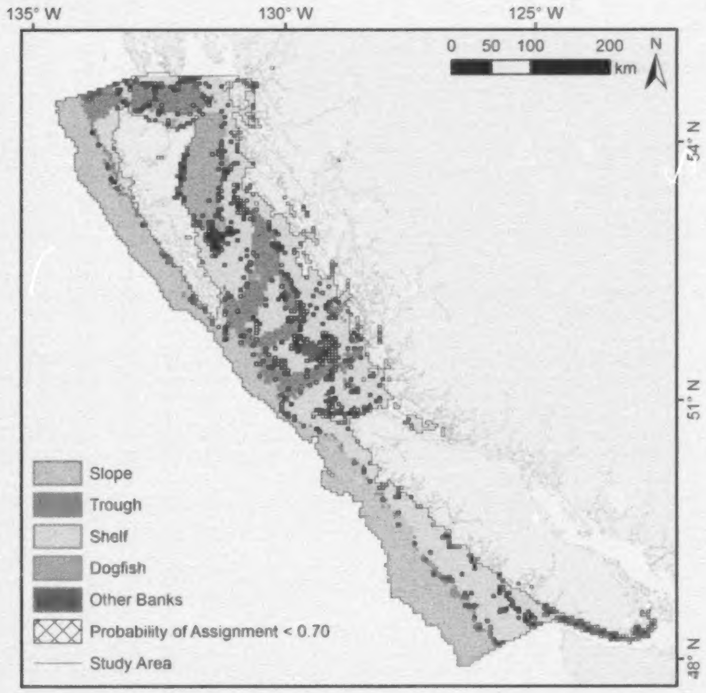


Figure 18. Level 4b (biophysical units) classification showing sites with low (< 0.70) probability of assignment from the random forest analysis. Probabilities are calculated as the proportion of overall trees in which a site is assigned to a class ("votes") of the number of trees assembled in the random forest (10,000).

3.4 LEVEL 4A AND B: BIOPHYSICAL UNITS DISCUSSION

A physiographic rule-based classification system was developed for Pacific Region during the past two decades, but its usefulness in MPA planning is unclear (see Appendix A for an analysis of BCMEC Ecosections biological relevance in comparison to our analysis). In this project, we focused our efforts on incorporating available biological information into PMECS as recommended by Robinson et al. (2015) and the national science advice on how to achieve representativity in MPA network design (DFO 2012b). We developed a two part analytical method that was used to populate Level 4 of PMECS to delineate major ecological units or large-scale habitat types and their associated species. In the first step we used a cluster analysis to examine patterns of species composition across space with the objective of identifying areas of similar species composition. We chose a hierarchical clustering methodology that allowed us to "cut" the dendrogram at different distances with increasing similarity between sites the closer a cut is to the branches of the dendrogram. Since the objective of this section was to identify and populate the level associated with depth-related processes at a large spatial extent (termed "bathome" in Last et al. 2005, Robinson et al. 2015), we were conservative in our cluster analysis, identifying clusters that are made up of sites with 35% similarity in composition (β_{sim} = 0.65) and β_{sim} = 0.55, or 45% similarity, to illustrate the hierarchical nature of the analysis. The second step of the analysis used the random forest approach to evaluate environmental correlates of the observed biological patterns, and to predict the ecological unit when the model is projected across the entire study area. The resulting predictive model allows gaps to be filled while also examining the underlying uncertainty in our model, as shown to be particularly apparent at boundaries between units.

The majority of benthic marine organisms are limited by some combination of depth, temperature, salinity, or substrate type (Roff and Zacharias 2011) and our results support this view with depth, summer salinity and temperature range as the top predictors of the biological assemblages at Level 4a and b in our study area. As hypothesized, the Biophysical units resulting from our analyses are strongly predicted by depth. For example, the Banks Biophysical unit, in addition to supporting a different biological assemblage, is shallower, fresher and warmer than the Shelf and Slope units. Similarly, the Slope unit contains a unique fauna and is deeper, colder with higher salinity than the Shelf and Banks Biophysical units. A strength of our approach is that we have the species composition data tied to the ecological unit so we not only have biological validation implicit in the classification, but also have a list of species that are associated with the ecological units. Both the Banks and Slope units had species with stronger associations than the Shelf unit, particularly the Slope unit where Tanner Crabs, Pacific Grenadier and Giant Grenadier were rarely found in other units indicating they are highly specific to Slope habitat.

The results of Level 4b found five broad-scale habitats in our study area with separate biological communities: Shelf, Troughs, Dogfish Bank, Other Banks, and the Slope. Interestingly, we found that the group of species on Dogfish Bank, the largest shallow bank in the region (Clarke and Jamieson 2006b), were distinct from other banks in the study area. This result supports the identification of Dogfish Bank as an Ecologically and Biologically Significant Area (EBSA, Clarke and Jamieson 2006b). An expert-driven process identified Dogfish Bank as an EBSA because it is the largest shallow area in the region, and is an important area of aggregation for marine birds and Dungeness Crab, and an important rearing habitat for flatfish and invertebrate larvae (Clarke and Jamieson 2006b, DFO 2012a). Interestingly, four species of flatfish (Pacific Sand Sole, Rock Sole, English Sole, and Butter Sole) were selected as indicator species for Dogfish Bank based on their high frequency in the cluster. All four flatfish species were also found in other Biophysical units, particularly Rock Sole in the Other Banks and Shelf units, however the higher frequency of flatfish in Dogfish Bank in comparison to other Biophysical units provides empirical evidence of its importance as flatfish habitat. Furthermore, although Dungeness Crab were found in low frequencies in other Biophysical units (2% of sites in Other Banks, and 2% of sites in Shelf), nearly half of the sites in the Dogfish Bank unit contained Dungeness Crab (49%), providing more evidence that this habitat is important for Dungeness crab aggregations, as outlined in its EBSA designation (Clarke and Jamieson 2006b). An added benefit of our community approach is the ability to provide a list of species associated with each ecological unit, information that is important to conservation planners and managers.

3.4.1 Model performance, uncertainty and field validation

Models of natural systems, including predictive ecological models like random forests inevitably include some degree of uncertainty. Uncertainty is not problematic *per se* as long as its effects on model projections are not ignored (Gould et al. 2014). However, many correlative models, in particular species distribution models (SDMs), are spatially projected without explicitly addressing uncertainty, thereby implying a confidence in model outputs that may be misleading (Beale and Lennon 2012, Wenger et al. 2013, Gould et al. 2014). Tulloch et al. (2015) stated that one of the most pervasive forms of uncertainty in data used to make conservation decisions is error associated with mapping of conservation features and while conservation planners should consider uncertainty associated with ecological data to make informed decisions, mapping error is rarely accommodated in the planning process. To better incorporate uncertainty into the planning process in the Pacific Region, we provided an uncertainty map that clearly highlights areas of lower confidence in model performance. It is important that the uncertainties in the PMECS classification approach be explicitly addressed to reduce the risk of adverse conservation or management outcomes (DFO 2013, Robinson et al. 2015). According to the performance metrics (Level 4a – large-scale biophysical unit: pseudo $R^2 = 0.93$ and AUC = 0.99; Level 4b – biophysical unit: pseudo $R^2 = 0.84$; AUC = 0.93), our model accurately classified sites using the environmental variables in the model suggesting that the structure of the communities is correlated with the environment; however, by examining the uncertainty, we were able to determine the spatial variability in our model accuracy.

We mapped the probability of assignment for the Level 4a and 4b of PMECS for each site to better understand the confidence level of the ecological unit assignment across the entire study area, and to capture the uncertainty in our model. The resultant maps (Figure 14, Figure 18) showed that areas of lower confidence aligned with the boundaries of the ecological units indicating transition zones. Model uncertainty in transition zones around edges of units is expected, and mapping it allows us to identify these potential transition zones. Transition zones are important features to consider and should be taken into account when delineating boundaries between biogeographic units (DFO 2009) and can be visualized with our approach. In contrast, when using a rule-based approach to classification that relies on univariate decision rules (examples include BCMEC, and our approach to the geomorphic units described in the next section), it can be difficult to identify a transition zone unless it is explicitly defined in the rules. The uncertainty map produced for the Biophysical units could be used as an input to MPA network design. For example, to ensure each Biophysical unit is represented in an MPA network, only sites with high confidence in the core of the biophysical units might be selected for inclusion, rather than areas with higher uncertainty. Transition zones could also be incorporated in the MPA network design process, given their ecological importance (Araujo 2002) and, in our study area, relatively higher species richness (see Appendix E).

Another advantage of mapping the uncertainty of the output classification is that it can be used to target field validation or groundtruthing on areas where the probability of assignment is low (< 70%). For example, we could target areas around Goose Bank in Queen Charlotte Sound, or the southern edge of Dogfish Bank to evaluate species presence/absence, with the understanding that these areas are potentially transition zones so species indicative of both the Shelf and the Other Banks biophysical units may be present. We also need to groundtruth the model in areas where it performed very well but no biological data have been collected. For example, the east coast of Haida Gwaii and Dogfish Bank, deeper sections of the Slope, and areas around Goose and Cook Banks on the shelf all have sizable biological sampling gaps. Sampling these areas, you would expect to have species present selected as Indicator species identified by the *Indval* analysis are expected to be present in samples from these areas.

In addition to collecting new data to groundtruth, we can also use existing data for groundtruthing that were not included in our original analysis. Several data sources are available that did not make it into our analysis because they did not meet our selection criteria. These sources could be used to test the classification output. For example, we could use spatially referenced presence data available for species associated with a specific biophysical unit to determine if those species are more prevalent in the biophysical units predicted by the model. In addition, expert opinion and First Nations knowledge could be used to complement or refine boundaries, particularly in areas of uncertainty.

3.4.2 Limitations of the approach

Ecological classifications often rely on physical and environmental "surrogates" to represent patterns in biodiversity. Surrogacy research in terms of marine habitat mapping is defined by Harris (2012b) as an "empirical method of determining which measured characteristics best describe the species assemblage in particular space and at a particular time." These "surrogates" then act as predictors for the occurrence of species assemblages in unexplored areas. Often classifications rely solely on physiographic data (AXYS Environmental Consulting

Ltd. 2001, Roff et al. 2003), whereas ideally classifications will include information on the distribution and abundance of species in association with the environment (Gregr et al. 2012). The choice of classification methodology is limited by the data available in the area to be classified. In areas where biological data is lacking, the best available abiotic data should be used to develop an ecological classification. Here, we combined biological and environmental surrogacy approaches to produce the large-scale biophysical unit and Biophysical unit classifications which correspond to Level 4a and 4b of PMECS, respectively.

The greatest limitation of using biological information in an ecological classification is the availability of appropriate data. The data need to be independent, representative and collected at a scale appropriate for the classification level. For this project, we were able to compile, error-check, and examine potential data sources and assess their use in a classification analysis. In addition, there was also a strong desire from managers in the area and previous work on PMECS to incorporate a biological component. Furthermore, national science advice on achieving representativity in MPA network design explicitly calls for the use of biological information when defining Bioregions if possible (DFO 2012b). Therefore, we developed an approach that uses both biological and environmental data to ensure biodiversity patterns are well represented in the ecological units at this scale.

The input species composition data that were used in the cluster analysis to define large-scale patterns in species assemblages included 174 species (96 demersal fish and 78 benthic invertebrates – see species list in Appendix D). Although this is just a fraction of the number of species that occur in the region, our analysis assumes that these species are representative of the benthic communities found in our study area. Biodiversity data often are not used in marine classifications is because they are lacking in the system of interest and because geophysical data often are available over broad geographic areas (Roff and Zacharias 2011, Gregr et al. 2012, Harris 2012a). We were able to collate suitable biological datasets that could be used in the development of a classification. A criticism related to the use of classifications built solely on species data (which ours is not) include: species are prone to local extinctions, invasion or extinction, but the community will persist and can be represented by recurrent physical processes (Roff and Zacharias 2011). Our analysis should be more robust than a single species approach as small changes in species co-occurrences resulting from local extinctions, range expansions, or novel introductions are less likely to alter whole communities than single species. Furthermore, because we incorporate the environmental variables into the model, it is the physical variables that define the boundaries of the biological communities rather than the species composition data. Another argument against a biological approach put forth by Roff and Zacharias (2011) is that physical habitats are far more temporally stable and ecologically fundamental than the communities they support. Physical habitats can be more stable than mobile species, although it depends on how physical habitat is defined. Attributes of physical habitats such as oxygen concentration, salinity, and temperature can be seasonally dynamic and ocean chemistry can change more rapidly than previously understood under global change (e.g., Curry et al. 2003, Orr et al. 2005, Deser et al. 2010).

We compared the biological distinctiveness of the BCMEC and the PMECS Biophysical unit layer (Level 4b) using non-metric multidimensional scaling (NMDS) and a PERMANOVA analysis and found that building the classification using our community approach results in more biologically distinct ecological units (see Appendix A). BCMEC, a physiographic classification, represents some biological patterns (particularly the Continental Slope Ecosection), but a methodology that combines both abiotic and biotic data is better at informing MPA network design in terms of ecological representativity. In addition, our approach allows for the prediction of the future distributions of known assemblages, by substituting current oceanographic data and models with projected conditions (e.g., D. Masson, DFO, is modelling changes in environmental variables over the next 100 years). Integration and analysis of a range of future scenarios (e.g., temperature rising 1 or 3ºC; freshwater input decreasing in certain areas) could help identify "buffer zones" around currently identified important areas, where we might expect changes in the boundaries of the ecological units to occur. Such an analysis could prioritize MPA sites within a network that are robust to environmental changes and have the highest probability of persisting in a changing ocean environment.

3.4.3 Conclusions

The classification methodology for Level 4 of PMECS provides a deeper understanding of biodiversity patterns in the Northern Shelf and the Southern Shelf Bioregions and a comprehensive analysis of the environmental correlates of those patterns. We maximized the use of available biological data on benthic species to develop a meso-scale classification delineating ecological units that represent distinct biological assemblages for use in MPA network planning. Our community ecological modeling output results in a classification of abiotic and biotic variables shaping biodiversity patterns in the Northern Shelf and Southern Shelf Bioregions in BC at two hierarchical scales. The resultant Biophysical units, Levels 4a and 4b, are defined by species assemblages and correlated environmental features and can be considered true ecological units because species-environment relationships and biotic interactions are incorporated into their delineations. We incorporated available biological data over a large spatial scale, despite the fact that the biological surveys were designed for another purpose. This study highlights the importance of maximising the use of biological data in marine conservation planning process. As more biological data become available they can be incorporated into this classification methodology.

4 LEVEL 5: GEOMORPHIC UNITS

4.1 GEOMORPHIC UNIT APPROACH

Level 5 "Geomorphic units" (100s of km scale) are used to identify ocean seabed features that are often associated with distinct biota (e.g., plateaus, ridges, seamounts, canyons).

Parks Canada in collaboration with the British Columbia Marine Conservation Analysis (BCMCA, BCMCA Project Team 2011) developed a coastwide application of a benthic habitat classification that was used in the Gwaii Haanas National Marine Conservation Area Reserve and Haida Heritage Site. This application adapts a benthic habitat model developed by The Nature Conservancy (TNC; Ferdaña 2006) and applied to marine ecoregional planning throughout the continental US (layer available at BCMCA, and see Robinson and Royle 2008 for application in Strait of Georgia). Following the TNC methodology, the benthic classes developed by Parks Canada/BCMCA combine three parameters: i) seascape features, ii) depth, and

iii) substrate, to identify areas of similar benthic characteristics. The Benthic Terrain Modeler (BTM) tool developed by NOAA Coastal Services was used to help generate four landscape features that describe the terrain of the seafloor (depressions, slopes, flats and ridges). These four landscape features were combined with four depth ranges and four types of substrate from an existing substrate layer created as input to the BC Marine Ecological Classification System (Ministry of Sustainable Resource Management (MSRM) Decision Support Services Branch 2002).The BCMCA habitat classification was not used directly in PMECS because of the issues already highlighted regarding the BCMEC Ecounits and their identification of substrate type (see section 1.3.1). Instead, here we use the Benthic Positioning Index and Benthic Terrain Modeler and other metrics derived from the highest resolution bathymetry layer available for NSB to define Geomorphological units for PMECS.

The BTM tool permits the creation of user-defined zone and/or structure classifications (Wright et al. 2005). These classifications are typically based on comparisons between broad and fine scale bathymetric position indices (BPI) and slope data (Lundblad et al. 2006), although some studies include predetermined features (i.e., named seamounts) or depth information (e.g., Ferdaña 2006). BPI is a measure of the elevation of a grid cell compared with the mean elevation of surrounding grid cells. Different scales are used to identify fine and broad scale features (Lundblad et al. 2006). The distance used to define broad versus fine scales are site specific and depend both on the overall topology of an area as well as the proposed use of the model. A recent study undertaken to identify large geomorphological features deeper than 30 m within the Canadian Pacific Exclusive Economic Zone (EEZ) used the BTM tool but with different user defined categories than those used by Lundblad et al. (2006) and others (Manson 2009, Appendix F Table F 1). Features were defined using broad scale BPI (500 m) and slope data and correlated to named seamounts and troughs.

Several other studies examining the Northeast Pacific Ocean have also used the BTM tool. Lanier et al. (2007) used BTM to create a map of surficial geological habitats off the coast of Oregon, USA. In this study, several analyses were utilized in order to capture the variability of the seascape over multiple scales (broad scale: 150 and 300 m; fine scale: 15, 25 and 50 m); geological features were defined as in Lundblad et al. (2006). Similarly, potential shallow water benthic habitats in a study site in northern California were mapped following the example of Lundblad et al. (2006), with broad scale data at 90 m and fine scale at 15 m (Erdey-Heydorn 2008). A recent study on habitat use of green sturgeon mapped BPI (scale: 250 m), including 6 user defined zones, and rugosity off the Oregon coast (Huff et al. 2011).

4.2 METHODOLOGY AND APPLICATION

4.2.1 Data

We used the BPI and BTM to analyze the 75 m resolution raster of bathymetry that was developed and made available by Robert Kung of Natural Resources Canada (NRCan) using data from the Canadian Hydrographic Service (CHS) to identify Geomorphic units. Natural Resource Map (NRM) contour lines, from the same CHS data used to develop the bathymetric raster, were also used for parts of this analysis. Prior to this work, these contour lines were merged to create a coast-wide dataset. The Northern Shelf Bioregion boundaries include the border with Alaska to the north, while the coastline to the east is defined by the high water line (GeoBC), and the base of the continental slope line (Ardron 2003). The Canadian Gazetteer of Undersea Feature Names was downloaded from the Government of Canada's Geogratis data repository.

4.2.2 Summary of methods

The following are the major steps in the analysis that produced geomorphic features:

1. Review literature on the use of the BPI to identify benthic zones and features and summarise parameters used;
2. Correct an error in the NRCan bathymetry raster;
3. Divide study area into smaller areas, such that each area contained features of roughly the same scale and had a more homogeneous seascape within. The features to be identified within each area could perhaps differ from adjoining areas;

4. For each of the smaller areas, use the BTM 3.0 beta tool (Wright et al. 2005) to produce classified maps. Through iterative trials, identify the most appropriate BPI scale parameters and appropriate thresholds to use in the classification dictionary;
5. Post-process the raster output to simplify boundaries, remove slivers, and join geomorphic units at the boundaries of the smaller analysis areas. Write metadata for final files; and
6. Attach names to features from the gazetteer of undersea feature names, as appropriate.

4.2.3 Bathymetry correction

A ‘hole’ was present in the southwest corner of the bathymetry file due to an error in the original contour dataset. One contour line had a depth attribute of 9000 m whereas based on neighbouring contours the correct value should have been 900 m. To rectify this error, the original contour file was updated to correct the erroneous contour line. The line was then converted to a point file using the vertices to point tool in ArcGIS. The Spatial Analyst IDW tool was then run with the same cell size as the CHS bathymetry raster to create a new bathymetry file. The Raster Calculator was run to convert the height values to negative numbers to match the CHS bathymetry file. Extract by Mask was used to extract only the area surrounding the bathymetry ‘hole’ and the resulting file was mosaicked to the original CHS raster.

4.2.4 Analysis area delineation

Initially we planned to divide the study area into two analysis areas – the continental shelf and the continental slope - because the scale of the features in each area differs significantly and the range of slope values (i.e., steepness of terrain) changes dramatically for the continental slope. Following a review of similar studies and a test run of the classification tool, it became apparent that we needed to separate out the coastal fjords, inlets, and channels into a third analysis area. The fjords themselves and any features to be delineated within them, are much smaller than the banks and troughs on the continental shelf.

In ArcGIS, the slope base was buffered by 2250 m (based on the BCMCA broad scale BPI) to extend the study area and ensure the analyses would not be affected by the edge. The polygon was joined to the EEZ using the Union tool. The resulting multipart polygon was exploded to individual polygons using the Multipart to Singlepart tool so that the offshore area could be removed. The Clip tool was then used to remove the terrestrial area based on the shoreline data (CHS high water line). The result was then dissolved to delineate the full study area, which is the marine area from the coastline to 2250 m past the base of the continental slope (Figure 19).

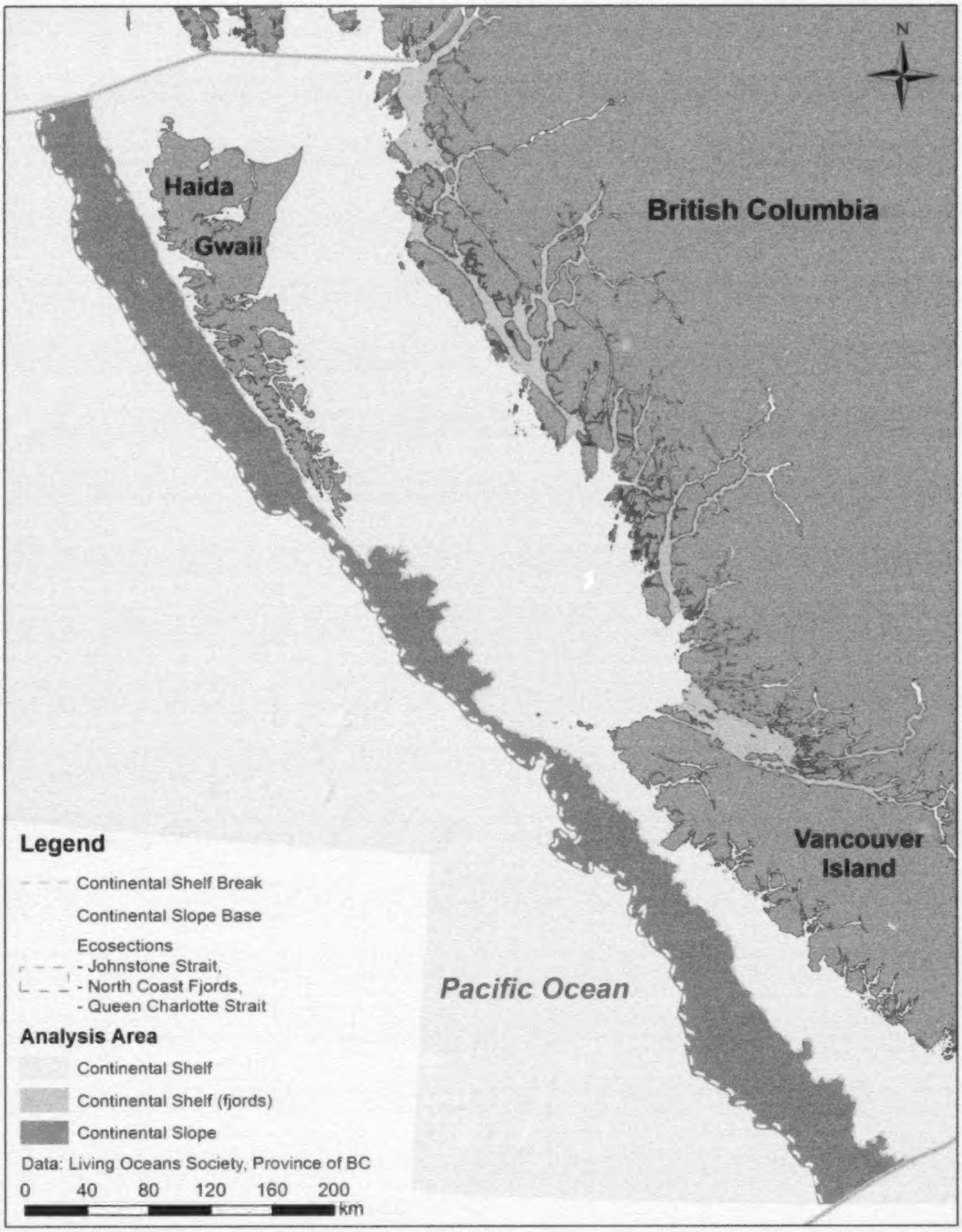

N
Haida
Gwaii
British Columbia
Vancouver
Island
Pacific Ocean
Legend
Continental Shelf Break
Continental Slope Base
Ecosections
- Johnstone Strait,
- North Coast Fjords,
- Queen Charlotte Strait
Analysis Area
Continental Shelf
Continental Shelf (fjords)
Continental Slope
Data: Living Oceans Society, Province of BC
0 40 80 120 160 200
km

Figure 19. Map of study area for geomorphic unit analysis.

4.2.4.1 Continental Slope

The offshore 200 m contour lines were selected, excluding inlets and closed loops, as the top of the continental slope and buffered by 2250 m. The resulting polygon was unioned with the full study area polygon and exploded using the Multipart to Singlepart tool. The shelf areas (east of the buffered 200 m contour) were removed. The result was dissolved to delineate the analysis area for the continental slope analyses.

4.2.4.2 Continental Shelf (with and without fjords)

The shelf area was delineated by tracing the offshore 200 m contour line, excluding inlets and closed loops. In order to capture the troughs within the shelf region, many of which are deeper than 200 m, the 500 m contour lines were used instead of the 200 m lines within Queen Charlotte Sound and Dixon Entrance. The contour lines were used to cut up the full study area polygon and areas offshore of the 200 and 500 m contours were deleted and the remaining polygons dissolved. The final polygon was cut again based on the southern boundary of the Northern Shelf Bioregion boundary and the area outside the bioregion was deleted. This analysis area overlaps purposefully with the continental slope analysis area to facilitate the delineation of geomorphic units along the border between the two analysis areas.

The Clip tool was used to extract the fjords from the main portion of shelf analysis area into a new continental shelf (fjords) feature class based on the North Coast Fjords, Queen Charlotte Strait, and Johnstone Strait ecosections (Ministry of Sustainable Resource Management (MSRM) Decision Support Services Branch 2002). The result was then erased from the shelf area file to extract the shelf area with no fjords into a continental shelf feature class.

The analysis area polygons were then used to extract a bathymetry raster for each analysis area. A slope raster was created for the whole Pacific Region and for the continental shelf without fjords using the BTM Slope tool.

4.2.5 Bathymetric Position Index parameters

The bathymetric position index (BPI) is a second order derivative and is derived from the first derivative (i.e., slope) of bathymetry. The BPI algorithm in the BTM tool compares each cell's depth to the mean depth of surrounding cells within a user defined annulus (donut shape) defined by inner radius and outer radius in units of cells. Using negative depth values results in negative BPI values for depressions, where the cell is lower than surrounding cells, positive values for crests or ridges, and near zero values for flats or areas of constant slope (Lundblad et al. 2006). The latter two can be distinguished using a threshold slope value.

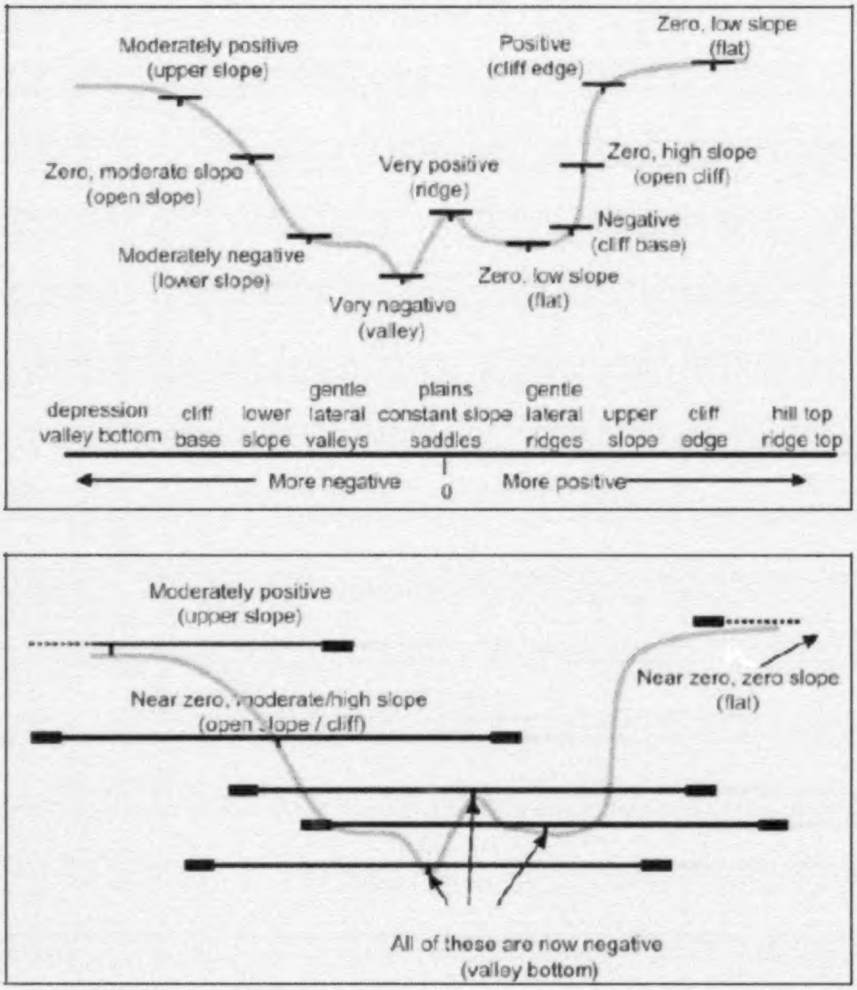

Moderately positive
(upper slope)
Positive
(cliff edge)
Zero, low slope
(flat)
Zero, moderate slope
(open slope)
Very positive
(ridge)
Zero, high slope
(open cliff)
Negative
(cliff base)
Moderately negative
(lower slope)
Very negative
(valley)
Zero, low slope
(flat)
depression
valley bottom
cliff
base
lower
slope
gentle
lateral
valleys
plains
constant slope
saddles
gentle
lateral
ridges
upper
slope
cliff
edge
hill top
ridge top
More negative
0
More positive
Moderately positive
(upper slope)
Near zero, moderate/high slope
(open slope / cliff)
Near zero, zero slope
(flat)
All of these are now negative
(valley bottom)

Figure 20. Top: with a small scale factor, subtle topographic features can be delineated. Bottom: a large scale factor results in broad scale topographic features (courtesy of Weiss 2001).

Using the broad and fine scale BPI tools in the BTM tool is the first step in classifying terrain into geomorphologic zones. While studies reported that scale of features should determine the scale factor of the BPI, none documented exactly how to determine appropriate values (Figure 20). We experimented with several different scale factors before developing our own guidance.

The BTM tool automatically calculates BPI rasters using inner and outer radii and the bathymetric raster as inputs. Each BPI raster was then standardized, using another BTM script, before it was ready to be input into the final classification tool.

4.2.6 Classification rules

The Classify Benthic Terrain script is the final script provided by the BTM tool. This script uses a classification table, standardized BPI rasters, slope, and sometimes depth rasters to delineate the geomorphologic features of interest. The classification table documents the rules that define each class or feature (i.e., Geomorphic units in this case). We developed a list of the type of features we anticipated delineating in each analysis area, then iteratively tested various classification tables until the features targeted for delineation best matched the analysis results.

The names of the Geomorphic units were based on the Greene et al. (2008) classification and were added to a field in the data attributes. This classification is composed of five codes describing megahabitat, sediment, meso- or macro-habitat, texture or biology and seafloor slope. The definitions of seafloor geomorphology found in Harris and Baker (2012b) were used to guide the naming procedure. Information was not available for sediment or texture and biology, so these codes were not included. Further, the meso- or macro-habitat level includes only two types of rises, mounds and ridges, so we included crest and wall as additional types to better describe the geomorphic features identified in this analysis.

4.3 GEOMORPHIC UNITS RESULTS

4.3.1 Bathymetric Position Index parameters

To determine appropriate BPI parameters, we listed features targeted for delineation and recorded measurements of a range of these features over the analysis area (Figure 19). The best match between features identified by the BPI and features of interest was obtained when the scale factor was approximately half the dimension of the features of interest. We also experimented with classifications using both broad and fine scale BPI layers and found that the best results were obtained when we developed the classification tables based on broad BPI values and slope (i.e., change in depth) values. Use of the fine scale BPI parameters generally added features that were too fine to be considered for the scale of desired Geomorphic units (e.g., wrinkles in the seascape that measured a few kilometers across).

Table 8. Features measured in each analysis area, range of sizes, and final scale factors chosen.

Analysis area	Features measured	Size range (width)	Broad scale BPI factor (m)	Inner radius (cells)	Outer radius (cells)
Fjords	Channels and inlets	2 to 10 km	3375	5	45
Continental Shelf	Troughs, banks	18 to 30, even 60 km	20025	27	267
Continental Shelf	Trough walls	1.5 to 5 km	Trough walls defined using slope, not BPI		
Continental slope	Canyons, valleys, ridges	2.3 to 7.7 km, Moresby Canyon - 13 km	3375	5	45

4.3.1.1 Continental slope

Features of interest on the continental slope that were measured in order to determine the range of scale factors to test included canyons, valleys and ridges and they ranged from 2.3 to 13 km wide. Scale factors ranging from 525 to 6525 m were tested. At the largest scale factor tested, 6525 m, Moresby Canyon, the widest canyon, was still not delineated into the depression class. The shape of this canyon is more consistently cupped across its width, so based on BPI values it is consistently identified as a slope feature. In addition, using the largest scale factor of 6525 m resulted in 'missing' some of the narrower named canyons off the West Coast of Vancouver Island, so a scale factor of 3375 m was chosen.

4.3.1.2 Continental shelf (without fjords)

Features on the continental shelf that were measured for delineation included wide troughs and banks, measuring anywhere from 18 to 30 km and even 60 km across. At the other end of the spectrum, trough walls were approximately 1.5 to 5 km wide. Scale factors ranging from 525 to 20025 m were tested. A relatively large scale factor of 20025 m was chosen, as it was best at delineating the troughs and banks. We found that the best way to delineate trough walls was to use a slope threshold in the classification table (see Classification rules).

4.3.1.3 Fjords

In the fjords analysis area, fjords and channels measured between 2 and 10 km across. Scale factors ranging from 525 to 12000 m were tested. The broad BPI scale factor chosen for the fjord analysis area was 3375 m. This provided a scale match to the continental slope analysis area and defined features that matched reasonably well at the join between the continental shelf analysis area and the fjords area.

4.3.2 Classification rules

4.3.2.1 Continental slope

Features to be delineated on the continental slope included ridges between canyons, depressions (i.e., canyon floors), sloping walls, and steeply sloping walls (Table 9). Ridges and depressions are generally defined as having extreme BPI values, greater than one standard deviation (SD) above or below the mean. One SD corresponds to a value of ± 100 in the standardized BPI raster. The 'flats' in this analysis area were areas of constant slope. This class, generally defined as within one SD from the mean, was broken into sloping and steeply sloping wall geomorphic units, using the slope threshold of 1.166, the mean slope value for the entire study area. The mean value was chosen because it divides the range of slope values into lower values corresponding to gently sloping walls and higher values corresponding to more steeply sloping walls. A similar method was used by Verfaillie et al. (2006).

Table 9 Classification table for continental slope analysis area.

Class	Zone	BroadBPI Lower	BroadBPI Upper	FineBPI Lower	FineBPI Upper	Slope Lower	Slope Upper	Depth Lower	Depth Upper
1	Ridge	100							
2	Canyon floor		-100						
3	Wall, sloping	-100	100				1.166		
4	Wall, steeply sloping	-100	100			1.166			

4.3.2.2 Continental shelf (without fjords)

Continental shelf features to be delineated included banks (possibly classed as flats or as crests depending on scale), troughs (which could be classed as flats or depressions again depending on scale), the slopes that determined the boundaries between banks and troughs, and areas of high rugosity or closely packed crests, depressions, and slopes. Overall in this analysis region, the features were of a larger scale, requiring a larger scale factor for the BPI, and the result of more subtle changes in depth. Crests and steeper trough bottoms (depression floors) were defined by extreme BPI values of ± more than one SD from the mean (Table 10). Delineating the more subtle troughs, named depressions, and banks, named mounds, required breaking the class of 'flat' (usually everything within 1 SD of the mean) into two classes, using the mean BPI value of 0.45 as the threshold. All flats with slope value greater than one were defined as sloping walls, as this value captured the sloped areas between the depressions and mounds. We did not use a rugosity raster to help delineate high rugosity Geomorphic units, although we do recommend further analysis to delineate the nearshore areas on the shelf that are currently closely packed crest, depression, and slope Geomorphic units (see Section 4.4).

Table 10 Classification table for continental shelf (without fjords) analysis area.

Class	Zone	BroadBPI Lower	BroadBPI Upper	FineBPI Lower	FineBPI Upper	Slope Lower	Slope Upper	Depth Lower	Depth Upper
1	Crest	100							
2	Depression floor		-100						
3	Depression	-100	0.45				1		
4	Mound	0.45	100				1		
5	Wall, sloping	-100	100			1			

4.3.2.3 Continental shelf (fjords)

In the fjords, inlets, and channels, we anticipated delineating steep slopes of fjord walls, depressions that were fjord and channel floors, flats at the heads of inlets, and possibly sill features close to where inlets or channels meet the continental shelf. Fjord walls (crests) and fjord or channel bottoms (depression floors) were defined by extreme BPI values of ± more than one SD from the mean (Table 11). We delineated depressions and mounds in the same manner as in the continental shelf (without fjords) analysis area. Finally, areas of constant but steep slope were defined as steeply sloping walls, similar to the continental slope analysis area, but using a threshold that matches the value used to define the slope break.

Table 11. Classification table for fjords analysis area.

Class	Zone	BroadBPI Lower	BroadBPI Upper	FineBPI Lower	FineBPI Upper	Slope Lower	Slope Upper	Depth Lower	Depth Upper
1	Crest	100							
2	Depression floor		-100						
3	Depression	-100	0.57				2.86		
4	Mound	0.57	100				2.86		
5	Wall, steeply sloping	-100	100			2.86			

4.3.3 Post-processing of classified Geomorphic units

The classified raster for each analysis area was exported to a feature class using the Raster to Polygon tool. The simplification option was selected during the export to produce smoother, less jagged polygons from the 75 m x 75 m raster grid. The simplification process should have little impact on the identification of Geomorphic units at a scale of 100s of kilometers. Because there were a significant number of small sliver polygons in the resulting files, the polygons were exploded using the Multipart to Singlepart tool and the distribution of polygon sizes were examined to identify sliver polygons. Slivers were defined as fragments found along the borders of the zones and the coastline and selected as polygons smaller than 1 km^2 for the continental slope, smaller than 0.5 km^2 for the continental shelf, and smaller than 0.2 km^2 for the fjords. The Eliminate tool was run to merge selected sliver polygons with the neighbouring polygon with which they share the longest border. This process was repeated a second time to ensure that obvious sliver polygons were removed. Because the continental slope and shelf analysis areas overlapped purposefully, to avoid artificially truncating geomorphic units at the boundary, a break point was required to ensure that geomorphic units did not overlap spatially. The 200 m isobath is not a very satisfactory proxy for the shelf break as the contour goes well into the shelf area where canyons intersect with the shelf (Figure 21). A shelf-break line feature was created following the methodology used previously to delineate a line for the base of the slope (Ardron 2003). The shelf-break was found to be delineated by a slope value of 2.86%. To create a smoother line for the shelf-break, the contour line closest to the break was traced. Because the bathymetry varies in space, the resulting line is a conglomeration of contours and does not represent a single isobath. The Geomorphic units for the continental slope and shelf were clipped based on the shelf break line so the features westward (offshore) of the line are from the continental slope analysis and features eastward (shoreward) of the line are from the continental shelf analysis. In a few areas, the continental shelf analysis area did not extend as far offshore as the shelf-break. The Geomorphic units for these areas were therefore taken from the continental slope analysis area results. The Eliminate tool was run on each analysis area to remove any slivers resulting from the clip using the same sliver parameters employed earlier.

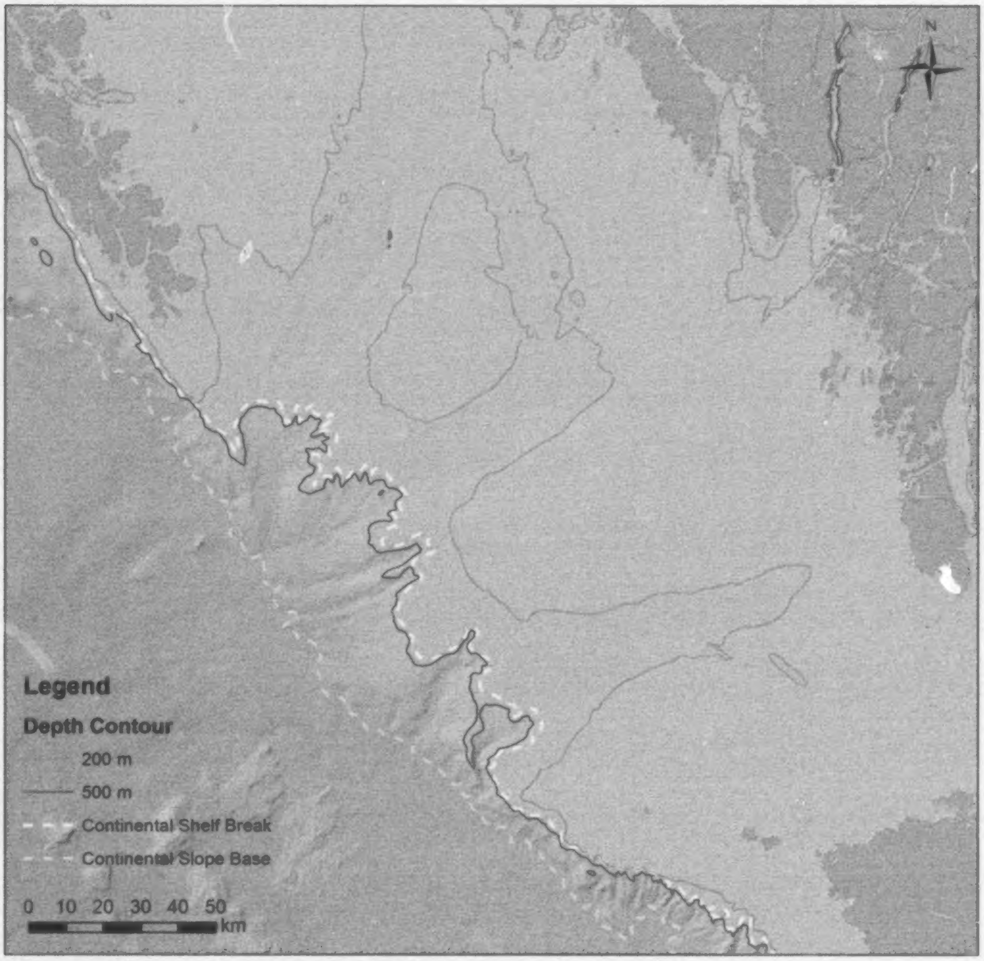


Figure 21. Continental slope base, newly defined continental shelf break with 200 and 500 m depth contours.

The three analysis area files were then merged into a single file. Because of the simplification procedure that was run when each raster file was converted to a polygon vector file, there were small gaps between the shelf and shelf (no fjords) analysis areas. Topology was created and edge-matching was performed to remove these gaps. The two continental shelf files were then merged using the Union tool and any resulting slivers smaller than 0.2 km^2 were eliminated. The final file for the continental shelf was then merged with the slope analysis area using the Merge tool. Topology was created for the final file and a few lingering gaps between the shelf and slope analysis areas were removed through edge-matching. Along the edge between the continental slope and shelf, there were a few long, narrow ridge and depression floor Geomorphic units identified between the continental slope steep slope Geomorphic units and the continental shelf slope Geomorphic units that were judged to be artifacts created by the presence of the analysis area edge. The ridge Geomorphic unit polygons were joined with the steeply sloping wall Geomorphic units from the continental slope analysis area while the depression floor Geomorphic unit polygons were joined with the sloping wall Geomorphic units from the continental shelf analysis area. Figure 22 illustrates the final geomorphic units.

The geomorphic units were compared to the point locations of gazetted undersea features and the canyon and trough features previously identified by Manson (2009). A field was added to the attribute table to record the names of associated gazetted features. The point locations of gazetted undersea features do not always match spatially with the features that were associated with that name. It is obvious looking at the bathymetry data that the point locations of gazetted features are not especially accurate. The attributes for the geomorphic spatial layer are described in Appendix F.

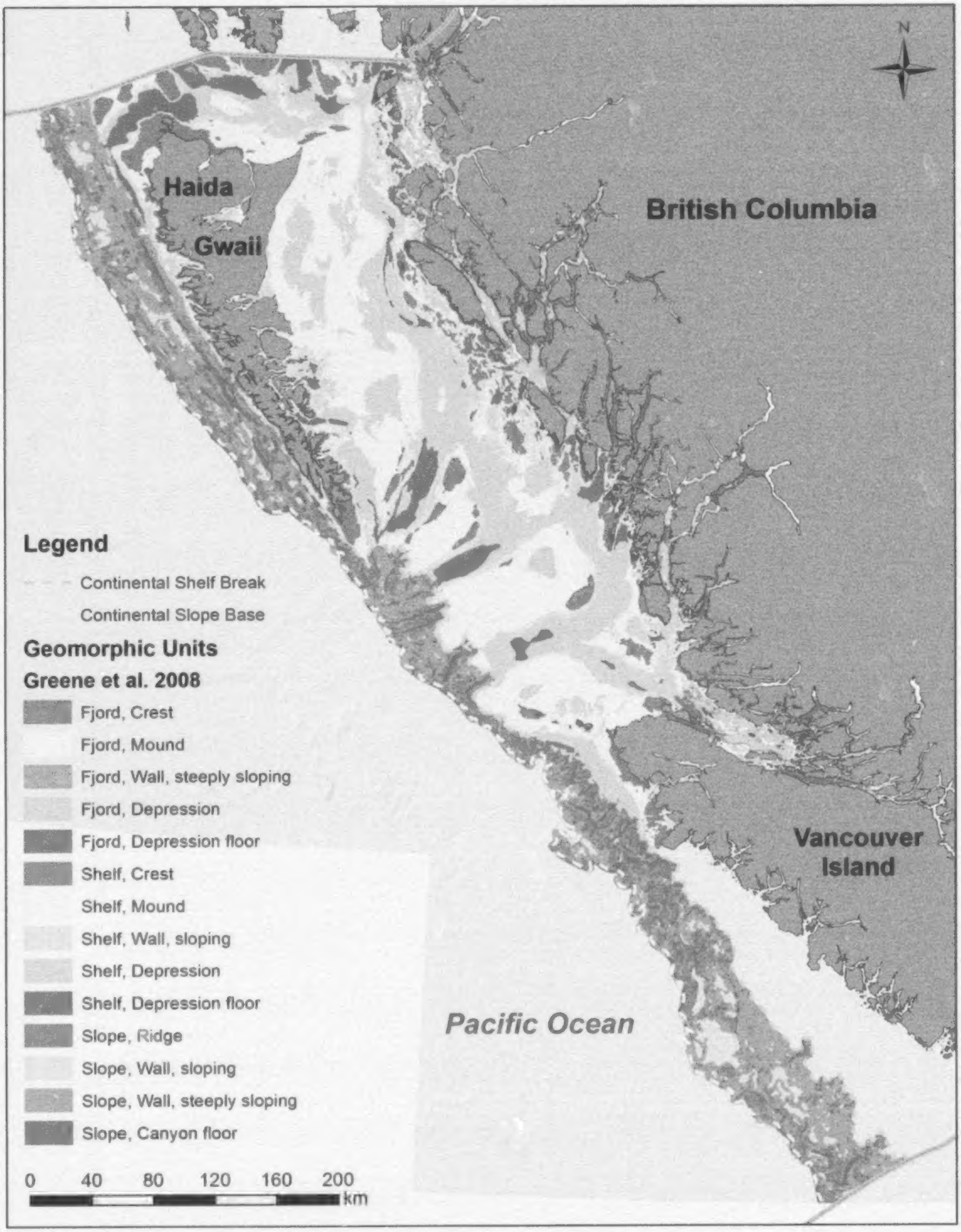

Haida
Gwaii
British Columbia
N
Vancouver Island
Pacific Ocean
Legend
Continental Shelf Break
Continental Slope Base
Geomorphic Units
Greene et al. 2008
Fjord, Crest
Fjord, Mound
Fjord, Wall, steeply sloping
Fjord, Depression
Fjord, Depression floor
Shelf, Crest
Shelf, Mound
Shelf, Wall, sloping
Shelf, Depression
Shelf, Depression floor
Slope, Ridge
Slope, Wall, sloping
Slope, Wall, steeply sloping
Slope, Canyon floor
0 40 80 120 160 200
km

Figure 22. Final geomorphic units for three analysis areas.

4.4 LEVEL 5: GEOMORPHIC UNITS: LESSONS, LIMITATIONS, AND FUTURE CONSIDERATIONS DISCUSSION

Two important lessons were learned from experimenting with the BTM tool. First, because the BPI is essentially a second order derivative, it is capable of delineating three shape classes only (i.e., consistent slope, ridge and depression). The tool is not designed to automatically identify complex features such as undersea canyons that may be defined as steep walls and a canyon bottom. This is true at both fine and broad scales. The tool does allows similar features to be defined at two different scales (e.g., a small scale ridge within a larger scale depression), but does not define complex features at one scale. Second, the scale of the target features dictates the best scale factor to employ, so advance identification of features targeted for delineation is helpful.

The CHS bathymetry raster was found to have data gaps within the fjords, some measuring almost 2 km in diameter, and a 'grid' artifact in offshore areas. Higher resolution bathymetry data with better coverage within the fjord areas would improve the results of future analyses. High resolution multibeam datasets are currently available for portions, though not the entirety, of the study area. Future analyses would also benefit from the use of the newly created shelf-break line to delineate analysis areas rather than the 200 m isobath.

One recommendation is to consider using the rugosity tool, also a part of the BTM tool kit, or an alternative method for calculating rugosity developed by Du Preez (2015), to delineate areas of high rugosity as distinct Geomorphic units. Currently, especially on the continental shelf, there are areas of tightly packed crest, depression, and slope Geomorphic units that are smaller than other units in this analysis area. These areas also correspond to known areas of high rugosity (e.g., as illustrated in the BCMCA atlas). If future work seeks associations between Geomorphic units and species distributions, then a continental shelf zone of high rugosity may be better suited for this than the current assemblage of many smaller Geomorphic units.

Future research should also assess the species assemblages associated with each geomorphic unit. The relationship between biological features and physiographic units has been discussed extensively in the literature (e.g., Harris 2012b) and there is substantial evidence that physical features are important surrogates for ecological structures, particularly at coarser levels of classification hierarchies (Roff and Taylor 2000, Beaman and Harris 2007, Greene et al. 2008). Abiotic factors have been utilized repeatedly as surrogates for biological communities based on the assumption that distinct Geomorphic units have distinct biota (e.g., Ferdaña 2006, Greene et al. 2008, Robinson and Royle 2008). Studies have confirmed a relationship between species patterns and physiographic features in the region (e.g., Zacharias and Roff 2001), but the correlation between species assemblages and geomorphic units specifically would benefit from more examination. Some of the biological data collated for the analyses of the Biophysical units may be appropriate for this work, depending on scale and locational accuracy.

5 LEVEL 6: BIOTOPES

"Biotopes" were defined in the conceptual PMECS framework as units at a spatial scale of 10s to 100 km that were nested within Geomorphic units and delineated into soft, hard or mixed substrate types. Substrate type, next to water depth, is among the most useful surrogates for delineating marine benthic communities (Harris and Baker 2012a). Once a substrate layer has been developed, the primary Biotope units can be delineated and described based on the substrate type and their associated biological communities. Over the past several years there has been a strong effort in the Pacific Region to develop a benthic habitat classification that can be used in the nearshore to identify and delineate sensitive habitats. For instance, Gregr et al. (2013) developed a spatial framework for representing nearshore ecosystems that delineates areas into primary and secondary Biotopes. These methods have been applied in the Strait of Georgia and Douglas Channel in the NSB. Given that this effort is ongoing in collaboration with departmental scientists and the methodology has been peer-reviewed, we recommend exploring the application of Gregr et al.'s (2013) bottom patch model to populate Level 6 of PMECS, as more data become available. Below we summarize the method from Gregr et al. (2013), with the understanding that some modifications will need to be made and certain data sources will vary depending on the depth and area being classified. The current methodology was developed to classify the bottom type from the high water line to 50 m in depth.

The bottom patch model, that classifies areas into Biotopes, was developed using data summarized in Table 12.

Table 12. Description of data sets used in Gregr et al. (2013) to classify nearshore benthic ecosystems (summarized from Gregr et al. (2013).

Type of Data	Description	Source
Two types of bathymetry maps	1) 75m raster	Produced by NRCan from CHS sounding data
	2) Depth polygons extracted from CHS electronic nautical charts previously compiled into a seamless polygon coverage	CHS
Six types of bottom type data including a variety of point and line features	1) ShoreZone (polyline) 2) Shellfish dive surveys (observational data - points) 3) Herring dive surveys (observational data – lines) 4) Parks Canada survey (observational data – points) 5) Hydrographic surveys (CHS – grab samples 6) Groundtruthing surveys (CHS – grab samples)	Province of BC, DFO, Parks Canada, and CHS
Tidal Energy	1) Point data	Mike Foreman

The methodology used to create the substrate classification includes five basic steps:

1. gather available substrate data;
2. assign a common classification;
3. create Thiessen polygons from each source;
4. combine polygons with the depth zones and the substrate layer; and

5. assign resulting bottom patches a type (BType) and a confidence (BConf) value based on the relationship between the patches and different data sources.

Figure 23 provides an example of the output of a bottom patch classification that may identify different Biotope units.

Nested within these putative Biotopes, if data are available, are smaller-scale, abiotic and/ or biotic sub-structural units of the seafloor (e.g. Hard-bedrock dominant, Hard-boulder dominant, Soft-sand/shell, Soft-mud, etc.). Gregr et al. (2013) added a secondary code to the primary classification when the information was available. For example, hard substrates can comprise either bedrock or large boulders. These secondary codes can be used to delineate Secondary biotopes.

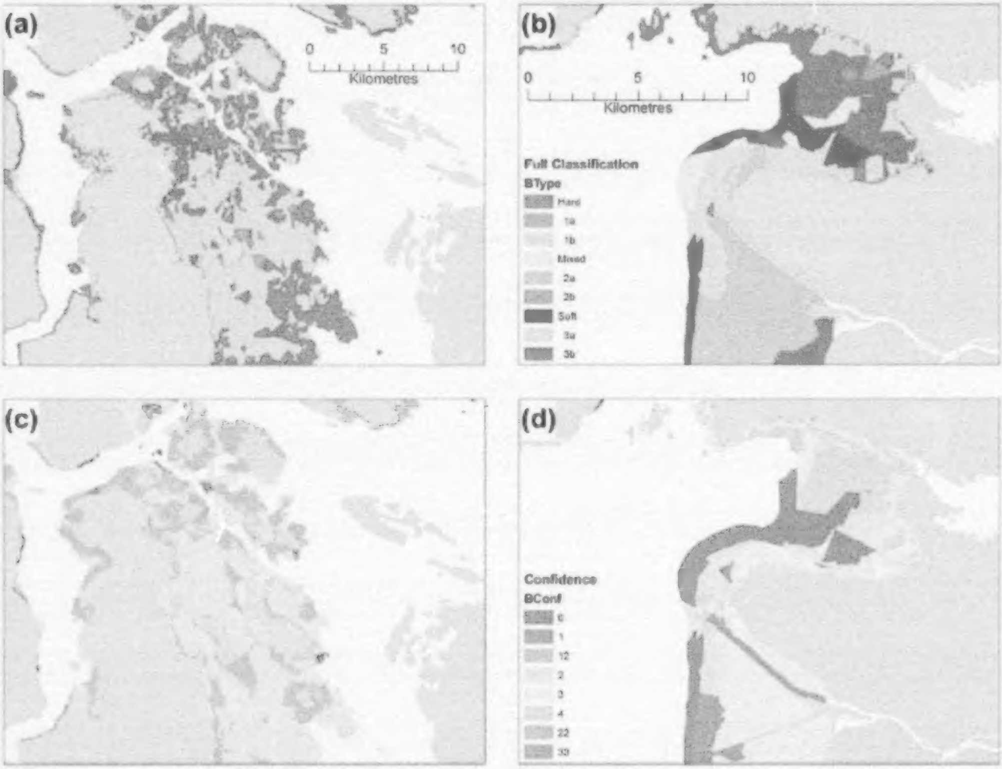


Figure 23. Example of Bottom Patch Classification: a and b show the bottom type while c and d show the confidence level of those classifications. Legends shown on b and d. Copied with permission from Gregr et al. (2013).

Gregr et al. (2013) evaluated the ecological performance of their bottom patches using observed shellfish distributions and found observations of Geoduck Clam (*Panopea generosa*), an infaunal species, and Red Sea Urchins (*Mesocentrotus franciscanus*), a species strongly associated with hard bottom, were significantly associated with soft and hard patches, respectively. The authors concluded that their model provides a reasonable approximation of primary bottom type but incorporating new data as they become available (e.g., acoustic backscatter analyses, multibeam sonar) will improve model fit. Currently, the bottom patch method has been applied to the entire north coast, Strait of Georgia and the east coast of Haida Gwaii (Gregr et al., University of British Columbia, Vancouver, B.C., personal communication,

2015) and represents the best available data to classify the substrate type layer at this scale. The methodology could be adapted to new data sources that are available below 50 m in depth. The application would need to be tested and validated for deeper regions prior to integration into PMECS.

6 LEVEL 7: BIOLOGICAL FACIES

Biological Facies are mappable units that act as surrogates for all levels of the hierarchical system at lower levels of PMECS (i.e., Micro-Assemblages, Species, and Genes, Table 2, Table 4; Last et al. 2010, Robinson et al. 2015). At a spatial extent of approximately 1 m to 1 km, they are identifiable by one or more indicator species (or group of species) that act as a surrogate for the broader biological assemblage to which they belong. Example of biological facies include biogenic habitat types such as Kelp forests, Eelgrass beds, and Glass Sponge reefs, all of which are associated with a specific ecological community. This is an important level within a hierarchical classification to populate for the management of biological diversity (Last et al. 2010, Robinson et al. 2015), but the full extent of Biological Facies across the Northern and Southern Shelf Bioregions is unknown and therefore data gaps are limiting their use as a management unit. In the Pacific Region, there are incomplete maps of the distribution of several Biological Facies including Glass Sponge reefs, Eelgrass beds, and Kelp forests that can be used in management and conservation planning. Additional survey data throughout the Bioregions (e.g. underwater imagery, benthic samples, aerial photography, multibeam surveys) would improve the capacity to populate this level of PMECS. In the meantime, species distribution models (SDMs) coupled with validation and groundtruthing could aid in developing predictions on the distribution of key Biological Facies.

Glass Sponge reefs are Biological Facies and known reefs were initially identified using multibeam surveys and then confirmed with underwater imagery (see Conway et al. 1991, Conway et al. 2005, Conway et al. 2007). Figure 24 shows a map of the extent of the Hecate Strait and Strait of Georgia Glass Sponge reefs with representative photographs of the reefs. The Sponge reef maps can be identified as Biological Facies units and can be incorporated into Level 7 of the PMECS framework. As new reefs are discovered or refined as boundaries are groundtruthed with further sampling, they can be incorporated into Level 7 of PMECS. A similar approach could be used to incorporate Eelgrass and Kelp forest distributions as Biological Facies in PMECS Level 7.

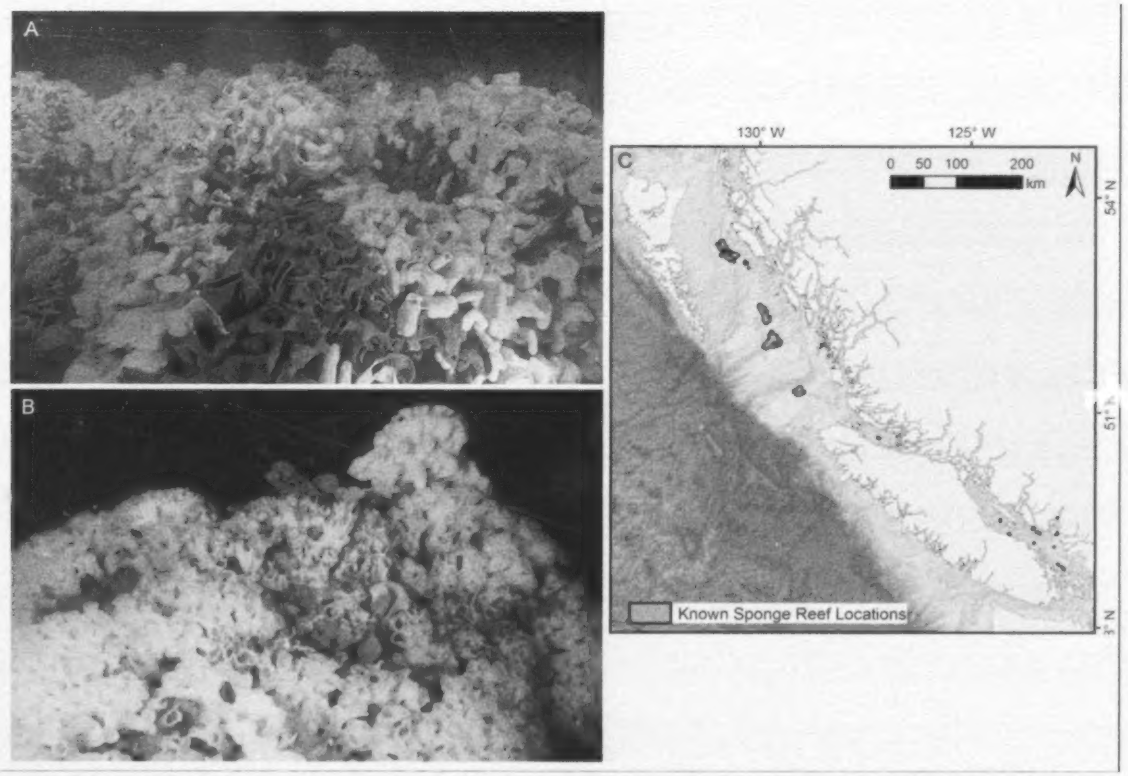


Figure 24. Glass sponge reefs as examples of biological facies. A) Image from Galiano Reef, Strait of Georgia. Courtesy J. Chu, S. Leys, and ROPOS. B) Image from Hecate Strait Reef. Courtesy DFO. C) Known extent of glass sponge reefs in Hecate Strait, Queen Charlotte Sound, and Strait of Georgia, BC

7 GENERAL DISCUSSION

The goal of this project was to progress the conceptual PMECS framework developed over the past several years (DFO 2009, 2013, Robinson et al. 2015) towards implementation for use in MPA network design. Mapping biodiversity patterns and understanding the environmental drivers of those patterns is a critical step in conservation planning, however, it is a challenging process, particularly in marine environments where data are limited and many information gaps exist. In this study, we compiled biological and environmental information from a broad range of sources into a central database and then selected appropriate layers to develop a marine ecological classification system that can be used to underpin marine spatial planning at biophysical and geomorphic levels. We developed an analytical approach to classify NSB and SSB of the Pacific Region to Level 4 of PMECS, a method that captures the broad-scale biological assemblages and the environmental variables correlated with them to produce mapped ecological units with a list of species highly associated with Biophysical units at two different scales (Level 4a and b). We also developed a rule-based classification using the broad-scale and fine-scale Benthic Positioning Index (BPI) to define Level 5, Geomorphic units, to capture the geomorphological features of the seabed. We developed methods to populate two levels of PMECS and have also suggested the integration of existing classification methods that can be used to populate the finer-scale levels of the PMECS framework (Biotopes and Biological Facies; Levels 6-7, Table 4) as more data become available. Specifically, we recommend adopting the Gregr et al. (2013) bottom patch model to classify benthic areas using available physical and biological substrate data as a method for identifying Biotopes (Level 6 of PMECS), assuming appropriate modifications can be implemented and validated for deep water and broader spatial scale. We also recommend collating available distribution data from key Biological Facies (e.g., Glass Sponge reefs, Eelgrasses, Kelp beds) and developing and evaluating models to predict their distributions in NSB and SSB. As outlined in Robinson et al. (2015), and similar to the Australian seabed regionalization (Last et al. 2010), the structure of PMECS ensures that larger-scale biodiversity patterns are captured at the top of the hierarchy (Levels 2–4) and at the bottom of the hierarchy in the more fine-scale levels (Levels 7–8), whereas abiotic variables can be used to populate the middle levels (Level 5–6).

7.1 CLIMATE CHANGE AND ECOLOGICAL CLASSIFICATION

Recent climate change studies show that patterns of warming of the upper layers of the world's oceans are significantly related to increases in greenhouse gases (Barnett 2005, Levitus et al. 2009). Results from a meta-analysis of climate impacts on marine species demonstrated a strong climate change fingerprint on marine life across the globe but also highlighted differences in responses among species, with 24% of species evaluated unexpectedly showing no response at all (Poloczanska et al. 2013). Hiddink et al. (2015) recently showed that different benthic invertebrates need to shift at different rates and directions to track the climate velocities of different temperature measures, and are therefore lagging behind most temperature measures. This lag effect could have implications on how benthic communities are likely to change and shift their distributions in the future. Climate change impacts and subsequent changes in community composition are important considerations for MPA network design and other marine spatial planning processes, but are challenging to capture in ecological classifications that require definition of boundaries. One of the advantages of the community approach taken in Level 4 of PMECS is that the random forest model can be used in a predictive framework. We can predict where these assemblages are likely to occur under future climate change scenarios to examine potential shifts. These predictions assume, however, that the assemblage-environment relationships will hold under a changing climate. Given the complexities of species interactions, and variation in individual species responses to climate change (as shown by Hiddink et al. 2015), the future plausibility of this assumption is unknown. However, the ability to use the classification in a predictive framework remains a strength of the Level 4 classification. In addition, the analytical nature (as opposed to rule-based) of the Level 4 methodology is also easily repeatable with new biological and environmental data to better understand how community composition is changing over time. Repeating the analyses periodically would provide a good understanding of how community composition is changing within the Biophysical units, but also if the Biophysical units themselves are shifting or disassembling.

7.2 HIERARCHICAL CLASSIFICATION SYSTEMS AND ISSUES OF SCALE

We developed a hierarchical ecological classification system that spans multiple spatial scales following national and regional guidance. At full implementation, PMECS will have broad application to varying management needs, from conservation planning, to fisheries management, and environmental impact assessment. However, the immediate need was to develop a hierarchical classification system to underpin MPA network design in the Pacific Region, with an initial focus on the Northern Shelf Bioregion (NSB).

An advantage of a hierarchical scheme for ecosystem-based management is that it provides context for the spatial information that is being considered, as well as a common reference framework for discussion and decision-making for managers (Harris 2012a). It is important to note that although a simple nested Log_{10} scale can be suggested in a hierarchical framework (i.e., 1000s of km, 100s of km, 10s of km, etc., see Table 2), in practice, the levels of the classification scheme cannot usually be specified in terms of a fixed-length scale, due to the natural patchiness and variability in seascapes. The nested scale of a hierarchical classification system is dependent on the area being classified and the scale of the PMECS level above. Despite its advantages, the choice of a hierarchical system is not without cost, as the more refined and detailed the hierarchical system, the greater the effort and resources needed to gather the necessary information to apply it to a particular area (see Table 4, Last et al. 2010, Robinson et al. 2015). Furthermore, if the information about the diversity of habitats is not collected at a scale that is relevant to its application or purpose, then the output risks being uninformative to managers as a decision-support tool (Harris 2012a).

Species–habitat relationships are strongly affected by the scale at which the dependent and independent variables are measured and studies suggest the scale of measurement can impact the strength and nature of observed species–environment relationships (Wiens 1989, Kotliar and Wiens 1990, Allen and Hoekstra 1991, Cushman and McGarigal 2004). In the Pacific Region, the scale of our analysis is limited by the scale at which data are collected. National science advice on achieving representativity in MPA networks explicitly states that classifications below a scale for which data are available should be avoided (DFO 2012b). Because of these issues, we initially focused on generating a broad-scale classification, both spatially (~1000s of km) and temporally (input data averaged/totaled over a 6–10 year time frame), to produce a classification that represents patterns of species assemblages and the main abiotic correlates shaping those patterns. In addition, we restricted our analysis to a 4 km grid cell to maximize the size of the sampling unit for our biological data (collected at a 2 km grid cell) while limiting the grid cell size to a resolution that is acceptable for use in species distribution models for conservation planning; Seo et al. (2009) has shown that higher resolution grid cells can overestimate species distributions, particularly for range-restricted species.

7.3 USE OF SPATIAL AUTOCORRELATION PATTERNS

Spatial autocorrelation, a pattern in which observations are related to one another by their geographic distance, is common in georeferenced ecological data (Legendre and Legendre 1998). The presence of spatial autocorrelation can create problems in species distribution modelling techniques (Lennon 2000, Dormann 2007, Crase et al. 2012) such as the random forest approach taken in this paper. Spatial autocorrelation in model residuals of SDMs violates the assumption of independent and identically distributed errors and can inflate type I errors (Legendre 1993, Kühn 2007), which can lead to the selection of unimportant explanatory variables and poorly estimated parameters in SDMs (Lennon 2000, Dormann 2007). There are several approaches to test for spatial autocorrelation in species distribution model residuals (reviewed by Keitt et al. 2002, Dormann et al. 2007) but our community modeling approach is different than the typical species distribution model making it more complicated to test for spatial autocorrelation. Firstly, we used geographic cohesiveness as one of several criteria for selecting a dissimilarity cut-off in our cluster analysis. In other words, we were looking for areas of similar species composition to map broad-scale benthic biological communities across geographic space, so spatial autocorrelation was inherent in our design (and could be considered a strength, see Gonzalez-Mirelis and Lindegarth 2012). Secondly, we have five classes (biological clusters) as the response variable in our model. Typically a SDM regression or machine-learning method is used to predict the distribution of one species using either presence/absence data or abundance data. To extract the residuals from our model, we would need to do it for each ecological unit prediction, and then test if the errors were correlated with geographic distance (using a Moran's I test), however given that our clusters are spatially coherent by design, some degree of spatial autocorrelation is likely.

A similar study that uses two types of classification trees, random forest and conditional inference forests, to predict benthic Biotope classes in a Swedish MPA, did not account for spatial autocorrelation in their predictive model (Gonzalez-Mirelis and Lindegarth 2012). The authors suggest that in distribution modelling, a model is fitted for the specific purpose of mapping its predictions, which involves using the mean, and the distribution of parameters are not often examined (as with this paper). Because only the variance of effects is largely affected by autocorrelation, Gonzalez-Mirelis and Lindegarth (2012) suggest spatial autocorrelation is less of a concern if the map is the output of interest (as opposed to the explanatory variables and parameter estimates). However, because spatial autocorrelation was not addressed in our model, we are unable to explicitly test the effects of the structuring processes of each ecological unit. In other words, although depth, summer salinity and temperature range are strong predictors of the biological clusters, we are limited in our interpretation regarding the strength of those correlative relationships.

7.4 INFORMATION GAPS

Information gaps limit the applications of our analyses at multiple levels within PMECS. At the Biophysical unit level, we were unable to uncover suitable multispecies datasets with comprehensive sampling in nearshore areas of the coast. Oceanographic data are also difficult to obtain for the shallower, narrow coastal areas. For these reasons, we removed sites that intersected with land from our analysis, resulting in no fjords, or inshore areas being classified at Level 4. Inshore areas are often significantly different than offshore areas due to differences in anthropogenic and naturally-induced pressures as well as differences in community structure and due to such differences, DFO (2012b) recommended these environments be considered separately when selecting the appropriate scale for incorporating representativity in the MPA network.

At the Geomorphic unit scale (Level 5), notable gaps in bathymetry data result in uncertainties associated with the delineation of physical features. As comprehensive high resolution bathymetry data are obtained (e.g. with multibeam sonar), features can be refined through re-analysis of these data and their derivatives (e.g. BPI, slope).

Classification at the Biotope scale (Levels 6) suffers similar issues with limited data resolution and lack of comprehensive and systematic data collection throughout the study area. However, the bottom patch methodology of Gregr et al. (2013) has the potential to be applied and refined as additional data become available (Table 3; Gregr et al. 2013).

Finally, to better define lower levels of the classification hierarchy, specifically Biological Facies, we need fine scale biological/physical surveys (10s m – 1 km). This includes fine scale biological surveys (dive surveys, underwater imagery, etc.) coupled with fine scale abiotic data (multibeam and backscatter, grab samples, CTD data) to first accurately map the known distributions of Biological Facies, but also to build strong ecological models that can be used to predict the presence of biogenic habitats in other areas.

7.5 SUMMARY AND FURTHER WORK

We collected data sources for a pelagic classification using the conceptual PMECS framework as a guide. There are several challenges with pelagic classifications, most notably the highly dynamic nature of the water column and movement patterns of pelagic species. To best deal with these challenges a focus on recurrent natural features as suggested by Roff and Zacharias (2011) may be the best approach. Although recurrent natural features like eddies, upwelling, and currents may shift regularly and have no fixed boundaries, their existence could be incorporated into a flexible classification system that adapts to these shifts. Additionally, examining a temperature-salinity diagram would be useful in identifying distinct water masses for a pelagic classification. Many key oceanographic features have already been identified as ecologically and biologically significant areas in the NSB and SSB bioregions (Clarke and Jamieson 2006a, 2006b).

In Table 13 below, we have recapped the recommendations from DFO (2013) and how they were addressed in our working paper to provide a summary of how PMECS development has progressed.

Table 13. Summary of recommendations from CSAS Science Advice Report (DFO 2013/065) and how these were addressed in our working paper.

SAR Recommendation	Implementation comments
1. Classification system must be hierarchical with multiple levels spanning bioregions to micro-communities (i.e., multiple and nested spatial scales).	PMECS was developed using hierarchical spatial scales but outputs are not perfectly nested due to natural patchiness.
2. Benthic and pelagic classification systems are needed.	Initial phase of PMECS focused on benthic classification however, data have been collected for pelagic classification as well.
3. PMECS needs to be created and evaluated with respect to management objectives at varying spatial scales.	Initial focus of PMECS is on MPA network design principles with an emphasis on "representativity". Classifications at biophysical and geomorphic units can be used to address representativity (and replication principles). Future classification at biotope and biological facies levels could address other MPA network design criteria, including integration of EBSAs (e.g., Glass Sponge reefs).
4. Application of a suite of tools to analyze and summarize biotic and abiotic data	Methodologies vary according to data type and spatial scale.
5. Identification of important data sources and gaps.	Key data sources identified and collated for Levels 4-5, and data gaps identified for Levels 4 - 7.
6. Uncertainty must be captured and documented (including the drivers of uncertainty) at each level in the framework.	Detailed metadata recorded and assumptions outlined, sources of uncertainty are identified and discussed.
7. Guidance is needed on how to address some known uncertainties/error in fundamental data sources (i.e., acoustic bathymetry and bottom type, water properties, abiotic and biotic surveys, and satellite, photographic and video imagery). This guidance will document the procedure for uncertainty documentation.	Uncertainty in underlying layers is outlined and uncertainty maps of PMECS Level 4 provided. A comprehensive uncertainty and/or sensitivity analyses of each abiotic layer used in the analysis is beyond the scope of this phase of PMECS development.
8. Guidance needs to be developed for the process for incorporating new data as it becomes available.	This paper describes the data and methodologies suitable and available to populate each layer of PMECS with available data. Once methodologies are reviewed, guidance for incorporating new data can be provided.

SAR Recommendation	Implementation comments
9. The performance of the PMECS prototype will be tested with existing data and evaluated with the appropriate metrics.	We test the performance of Level 4 using standard performance metrics. For the geomorphic units, we compared results to previous results by Manson (2009) and gazetted place names.
10. PMECS should make use of DFO's ongoing climate change research projections and trends.	Opportunities for doing so are discussed in the working paper.
11. A parallel pilot program should be developed to evaluate the performance of different methods using the same datasets to provide guidance on model choices to PMECS users.	Beyond scope of this stage in PMECS development.
12. Collaboration is a critical component of PMECS development. PMECS development is analyses heavy and requires multiple data sources, technical expertise and data sharing. It depends on numerous partners.	Established core science team with members from the province, DFO, and FN and consultation with academics. Also solicited feedback from national working group. Compiled 600 data layers from multiple sources.
13. There is a recognized gap in the coordination of GIS capacity for geospatial planning and management internally within DFO and externally with other agencies/partners. Standards for data collection, storage and sharing will be needed as implementation of PMECS proceeds.	Central PMECS database on server, attempts to coordinate data requests with Oceans. Will be available for input into Federal Geospatial Platform (FGP) when established and finalized.
14. The development of PMECS, following the guidance and advice given during the PMECS Regional Peer Review, will result in progress towards fulfilling national commitments to ensure that adequate MPA network design is achieved in the Pacific Region.	Initial focus on MPA fulfilling network design principles.

8 ACKNOWLEDGEMENTS

We are grateful for feedback and discussion from several people during the development of this working paper including: Ed Gregr, Greig Oldford, Chris McDougall, Jim Boutillier, Michelle Greenlaw, Krista Royle, Peter Lawton, Marty King, Marc Ouellette, Natalie Ban and Miriam O. We also would like to thank Kate Rutherford, Leslie Barton, Pramod Thupaki, Shelee Hamilton, Jason Dunham, Martin Huang, Peter Wills, Melissa Frey, Gavin Hanke and many others who provided access to and answered questions about our data sources. We are also grateful for the feedback and reviews provided from several CSAS participants including those already mentioned above but in addition: Cliff Robinson, John Harper, John Roff, Lynn Lee, Tetjana Ross, Charles Hannah, John Holmes, Marilyn Hargreaves, and Joanne Lessard.

9 DEFINITIONS

9.1 GLOSSARY

Biodiversity – The full range of variety and variability within and among living organisms and the ecological complexes of which they are a part (Canada – British Columbia Marine Protected Area Network Strategy 2014).

Biotope – Discrete taxonomic assemblages characterized by associated substrate and environmental factors (see Table 4 and Section 5).

Habitat – The physical and biological conditions that make up the environment where organisms live.

Hierarchy – A classification system where sets of objects or concepts are organized by rank or level. In a perfectly nested hierarchy, each set contains a subset in the level beneath it, and each subset belongs to only one set above it. Geographic hierarchical classifications contain different conceptual levels that are nested by geographic location, spatial scale, or both. The PMECS classification is nested by spatial scale, with smaller classification units falling within larger classification units, but is not necessarily perfectly nested by area. That is, the boundaries of smaller units may overlap with multiple larger units.

Representativity – Relatively intact, naturally functioning examples of the full range of ecosystems and habitat diversity found within a given planning area (Canada – British Columbia Marine Protected Area Network Strategy 2014).

Rugosity – A measure of terrain complexity or roughness. Generally calculated as the ratio between the contoured distance/area between two points (i.e., over the surface) and the corresponding planar distance/area (Du Preez 2015).

Species distribution – The spatial or geographic arrangement of a biological taxon.

Species richness – The number of species present in a defined area.

9.2 ACRONYMS

AquaMODIS – Moderate Resolution Imaging Spectroradiometer, aboard NASA's Aqua satellite.

AUC – Area under the receiver operating curve; a metric used to measure predictive modelling accuracy.

BCMCA – British Columbia Marine Conservation Analysis.

BCMEC – British Columbia Marine Ecological Classification System.

BPI – Bathymetric Position Index

BTM – Benthic Terrain Modeler, a set of tools in ArcGIS for analyzing benthic terrain.

ChlA – Chlorophyll A

CHS – Canadian Hydrographic Service.

CTD – An instrument used to measure water conductivity, temperature, depth, and other parameters.

EBSA – Ecologically and Biologically Significant Areas

EEZ – Exclusive Economic Zone ; the area, extending from a state's coast to 200 nautical miles offshore, over which states have exclusive jurisdiction and rights over exploration and exploitation of marine resources.

GF Bio – Groundfish biological survey database; DFO data holdings for scientific research regarding groundfish fisheries.

IDW – Inverse Distance Weighting; a method of interpolating data from point values.

ITIS – Integrated Taxonomic Information System.

NASA – National Aeronautics and Space Administration.

NOAA – National Oceanographic and Atmospheric Administration.

NRCan – Natural Resources Canada.

RBCM – Royal British Columbia Museum.

SDM – Species distribution model; statistical tools used to predict the occurrence or abundance of organisms in a given area.

SF Bio – Shellfish biological survey data holdings; DFO data holdings for scientific research regarding shellfish fisheries.

SF log – Shellfish commercial log data holdings; DFO data holdings of records from commercial shellfish fisheries.

WoRMS – World Register of Marine Species

10 LITERATURE CITED

Allen, L.G., and Pondella, D.J., II. 2006. Ecological Classification. *In* The Ecology of Marine Fishes: California and Adjacent Waters. *Edited by* L.G. Allen, D.J. Pondella, II and M.H. Horn. University of California Press, Berkeley and Los Angeles, California.

Allen, M.J., and Smith, G.B. 1988. Atlas and Zoogeography of Common Fishes in the Bering Sea and Northeastern Pacific. NOAA Technical Report NMFS 66. National Marine Fisheries Service, National Oceanic & Atmospheric Administration.

Allen, T.F.H., and Hoekstra, T.W. 1991. The role of heterogeneity in scaling of ecological systems under analysis. *In* Ecological heterogeneity. Ecological studies 86. *Edited by* J. Kolasa and S. Pickett. Springer-Verlag, New York. pp. 47-68.

Ardron, J.A. 2003. Data Description: Base of Shelf Slope. Living Oceans Society.

Araujo, M.B. 2002. Biodiversity hotspots and zones of ecological transition. Conservation Biology 16:1662-1663.

AXYS Environmental Consulting Ltd. 2000. British Columbia Marine Ecological Classification Update – Method Options. Prepared for Land Use Coordination Office, Government of British Columbia.

AXYS Environmental Consulting Ltd. 2001. British Columbia Marine Ecological Classification Update. Ministry of Sustainable Resource Management Decision Support Services.

Bailey, R.G., Zoltai, S.C., and Wiken, E.B. 1985. Ecological regionalization in Canada and the United States. Geoforum 16: 265-275

Ban, N.C. 2009. Minimum data requirements for designing a set of marine protected areas, using commonly available abiotic and biotic datasets. Biodiv. Cons. 18(7): 1829-1845

Barnett, T.P. 2005. Penetration of human-induced warming into the world's oceans. Science 309(5732): 284-287

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. Global. Ecol. Biogeog. 19(1): 134-143

Beale, C.M., and Lennon, J.J. 2012. Incorporating uncertainty in predictive species distribution modelling. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367: 247-258

Beaman, R.J., and Harris, P.T. 2007. Geophysical Variables as Predictors of Megabenthos Assemblages from the Northern Great Barrier Reef, Australia. *In* Mapping the Seafloor for Habitat Characterization. *Edited by* B.J. Todd and H.G. Green. Geological Association of Canada Special Paper 47. pp. 247-263.

Bett, B.J., and Narayanaswamy, B.E. 2014. Genera as proxies for species α- and β-diversity: tested across a deep-water Atlantic-Arctic boundary. Mar. Ecol. 35(4): 436-444

Bredin, K.A., Gerriets, S.H., and Van Guelpen, L. 2001. Distribution of rare, endangered, and keystone marine vertebrate species in Bay of Fundy seascapes. *In* Health of the Bay of Fundy: Assessing Key Issues. Proceedings of the 5th Bay of Fundy Science Workshop and Coastal Forum "Taking the Pulse of the Bay", Wolfville, Nova Scotia, May 13-16, 2002. *Edited by* P.G. Wells, G.R. Daborn, J.A. Percy, J. Harvey and S.J. Rolston. Environment Canada - Atlantic Region, Occaisonal Report No. 21, Dartmouth, NS and Sackville, NB. pp. 83-98.

Breiman, L. 2001. Random Forests. Mach. L. 45: 5–32

British Columbia Marine Conservation Analysis (BCMCA) Project Team. 2011. Marine Atlas of Pacific Canada: A Product of the British Columbia Marine Conservation Analysis. (Accessed 15 March 2016)

Canada – British Columbia Marine Protected Area Network Strategy. 2014. (Accessed 15 March 2016)

Clarke, C.L., and Jamieson, G.S. 2006a. Identification of Ecologically and Biologically Significant Areas in the Pacific North Coast Integrated Management Area: Phase I - Identification of Important Areas. Can. Tech. Rep. Fish. Aquat. Sci. 2678

Clarke, C.L., and Jamieson, G.S. 2006b. Identification of Ecologically and Biologically Significant Areas in the Pacific North Coast Integrated Management Area: Phase II – Final Report. Can. Tech. Rep. Fish. Aquat. Sci. 2686

Conway, K.W., Barrie, J.V., Austin, W.C., and Luternauer, J.L. 1991. Holocene sponge bioherms on the western Canadian continental shelf. Cont. Shelf Res. 11: 771–790

Conway, K.W., Barrie, J.V., and Krautter, M. 2005. Geomorphology of unique reefs on the western Canadian shelf: sponge reefs mapped by multibeam bathymetry. Geo-Mar. Lett. 25(4): 205-213

Conway, K.W., Barrie, J.V., Hill, P.R., Austin, W.C., and Picard, K. 2007. Mapping sensitive benthic habitats in the Strait of Georgia, coastal British Columbia: deep-water sponge and coral reefs. Geological Survey of Canada, Current Research 2007-A2

Crase, B., Liedloff, A.C., and Wintle, B.A. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35(10): 897-888

Curry, R., Dickson, B., and Yashayaev, I. 2003. A change in the freshwater balance of the Atlantic Ocean over the past four decades. Nature 426: 826-829

Cushman, S.A., and McGarigal, K. 2004. Patterns in the species environment relationship depend on both scale and choice of response variables. Oikos 105: 117-124

Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. 2007. Random forests for classification in ecology. Ecology 88: 2783-2792

Deser, C., Phillips, A.S., and Alexander, M.A. 2010. Twentieth century tropical sea surface temperature trends revisited. Geophys. Res. Lett. 37(10)

DFO. 2009. Development of a Framework and Principles for the Biogeographic Classification of Canadian Marine Areas. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2009/056.

DFO. 2012a. Evaluation of proposed Ecologically and Biologically Significant Areas in marine waters of British Columbia. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2012/075.

DFO. 2012b. Science Guidance on how to Achieve Representativity in the Design of Marine Protected Area Networks. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2012/083.

DFO. 2013. Key elements in the development of a hierarchical marine ecological classification system to support ecosystem approaches to management in Pacific Canada. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2013/065

Dormann, C.F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global. Ecol. Biogeog. 16(2): 129-138

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5): 609-628

Druehl, L. 2000. Pacific Seaweeds. Harbour Publishing, Madeira Park, BC.

Du Preez, C. 2015. A new arc–chord ratio (ACR) rugosity index for quantifying three-dimensional landscape structural complexity. Landscape Ecol. 30(1): 181-192

Dufrêne, M., and Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible symmetrical approach. Ecol. Monog. 67(3): 345-366

Erdey-Heydorn, M.D. 2008. An ArcGIS seabed characterization toolbox developed for investigating benthic habitats. Mar. Geod. 31(4): 318-358

Fargo, J.J. 2012. Dynamics of Marine Fish Assemblages in Hecate Strait, British Columbia. Can. Tech. Rep. Fish. Aquat. Sci. 2996

Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27: 861-874

Fenberg, P.B., Menge, B.A., Raimondi, P.T., and Rivadeneira, M.M. 2015. Biogeographic structure of the northeastern Pacific rocky intertidal: the role of upwelling and dispersal to drive patterns. Ecography 38(1): 83-95

Ferdaña, Z. 2006. Case study: Pacific Northwest Coast (PNWC) ecoregion offshore classification methodology. *In* Ecoregional Assessment and Biodiversity Vision Toolbox *Edited by* J. Higgins and R. Esselman. The Nature Conservancy, Arlington, VA.

Foreman, M.G.G., Crawford, W.R., Cherniawsky, J.Y., and Galbraith, J. 2008. Dynamic ocean topography for the northeast Pacific and its continental margins. Geophys. Res. Lett. 35(22): 5 pp

Franklin, J. 2009. Mapping species distributions - spatial inference and prediction. Cambridge University Press, New York, USA.

Gauch, H.G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Greenlaw, M.E., Smith, K., Martin, R. 2015. A subtidal ecological classification system to represent species diversity and distribution patterns in the Maritimes Region. DFO Can. Sci. Advis. Sec.Res Doc. In press.

Gonzalez-Mirelis, G., and Lindegarth, M. 2012. Predicting the distribution of out-of-reach biotopes with decision trees in a Swedish marine protected area. Ecol. Appl. 22(8): 2248-2264

Gould, S.F., Beeton, N.J., Harris, R.M., Hutchinson, M.F., Lechner, A.M., Porfirio, L.L., and Mackey, B.G. 2014. A tool for simulating and communicating uncertainty when modelling species distributions under future climates. Ecol. Evol. 4(24): 4798-4811

Greene, H.G., O'Connell, V., Brylinsky, C.K., and Reynolds, J.R. 2008. Marine Benthic Habitat Classification: What's Best for Alaska? *In* Marine Habitat Mapping Technology for Alaska. *Edited by* J.R. Reynolds and H.G. Greene. Alaska Sea Grant College Program, University of Alaska Fairbanks, Fairbanks, Alaska. pp. 169-184.

Gregg, W.W., and Casey, N.W. 2007. Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. Remote Sens. Environ. 111: 25-35

Gregr, E.J. 2012. Metadata, bc_eez_100m. (Accessed 15 March 2016)

Gregr, E.J., Ahrens, A.L., and Perry, I.R. 2012. Reconciling classifications of ecologically and biologically significant areas in the world's oceans. Mar. Pol. 36(3): 716-726

Gregr, E.J., Lessard, J., and Harper, J. 2013. A spatial framework for representing nearshore ecosystems. Prog. Oceanog. 115: 189-201

Harper, J.R., Christian, J., Cross, W.E., Firth, R., Searing, G., and Thompson, D. 1993. A classification of the marine regions of Canada. Final Report to Environment Canada, Vancouver, B.C.

Harris, P.T. 2012a. Biogeography, Benthic Ecology, and Habitat Classification System. *In* Seafloor Geomorphology as Benthic Habitat. *Edited by* P.T. Harris and E.K. Baker. Elsevier, San Francisco. pp. 61-87.

Harris, P.T., and Baker, E.K. 2012a. GeoHab Atlas of Seafloor Geomorphic Features and Benthic Habitats: Synthesis and Lessons Learned. *In* Seafloor Geomorphology as Benthic Habitat. *Edited by* P.T. Harris and E.K. Baker. Elsevier, San Francisco. pp. 871-890.

Harris, P.T. 2012b. Surrogacy. *In* Seafloor Geomorphology as Benthic Habitat. *Edited by* P.T. Harris and E.K. Baker. Elsevier, San Francisco. pp. 93-102.

Harris, P.T., and Baker, E.K. 2012b. Glossary. *In* Seafloor Geomorphology as Benthic Habitat. *Edited by* P.T. Harris and E.K. Baker. Elsevier, San Francisco. pp. 891-900.

Hernandez, F.J., Jr., Carassou, L., Graham, W.M., and Powers, S.P. 2013. Evaluation of the taxonomic sufficiency approach for ichthyoplankton community analysis. Mar. Ecol. Prog. Ser. 491: 77-90

Hiddink, J.G., Burrows, M.T., and Garcia Molinos, J. 2015. Temperature tracking by North Sea benthic invertebrates in response to climate change. Glob. Chang. Biol. 21(1): 117-129

Howes, D., Harper, J.R., and Owens, E.H. 1994. Physical shore-zone mapping system for British Columbia. Resource Inventory Committee (RIC) Report by the Coastal Task Force, RIC Secretariat, Victoria, BC. (Accessed 15 March 2016)

Huff, D.D., Lindley, S.T., Rankin, P.S., and Mora, E.A. 2011. Green sturgeon physical habitat use in the coastal Pacific Ocean. PLoS One 6(9): e25156

Hurlbert, A.H., and Jetz, W. 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proc. Natl. Acad. Sci. 104(33): 13384–13389

Johannessen, D., Haggarty, D., and Pringle, J. 2004. Boundary Definition for the Central Coast Integrated Management Area. DFO Can. Sci. Advis. Sec. Res. Doc. 2004/050. iv + 58 p.

Keitt, T.H., Bjørnstad, O.N., Dixon, P.M., and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. Ecography 25(5): 616-625

Koleff, P., Gaston, K.J., and Lennon, J.J. 2003. Measuring beta diversity for presence–absence data. J. Anim. Ecol. 72: 367-382

Kostylev, V.E., Todd, B.J., Longva, O., and Valentine, P.C. 2005. Characterization of Benthic Habitat on Northeastern Georges Bank, Canada. Am. Fish. Soc. Symp. 41: 141-152

Kotliar, N.B., and Wiens, J.A. 1990. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. Oikos 59: 253-260

Kreft, H., and Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. J. Biogeog. 37(11): 2029-2053

Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. Divers. Distrib. 13(1): 66-69

Lanier, A., Romsos, C., and Goldfinger, C. 2007. Seafloor habitat mapping on the Oregon continental margin: a spatially nested GIS approach to mapping scale, mapping methods, and accuracy quantification. Mar. Geod. 30(1-2): 51-76

Last, P., Lyne, V., Yearsley, G., Gledhill, D., Gomon, M., Rees, T., and White, W. 2005. Validation of national demersal fish datasets for the regionalisation of the Australian continental slope and outer shelf (>40 m depth). National Oceans Office, Department of Environment and Heritage and CSIRO Marine Research, Australia, Hobart, Tasmania.

Last, P.R., Lyne, V.D., Williams, A., Davies, C.R., Butler, A.J., and Yearsley, G.K. 2010. A hierarchical framework for classifying seabed biodiversity with application to planning and managing Australia's marine biological resources. Biol. Cons. 143(7): 1675-1686

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74(6): 1659-1673

Legendre, P., and Legendre, L. 2012. Cluster Analysis. *In* Numerical Ecology, 3rd ed. Developments in Environmental Modelling Vol. 24. Elsevier, Amsterdam.

Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. Ecography 23(1): 101-113

Levings, C.D., and Jamieson, G.S. 1999. An evaluation on criteria for creating MPAs in the Pacific Region: A proposed semi-quantitative scheme.. DFO Can. Sci. Advis. Sec. Res. Doc. 1999/210. 30 p.

Levitus, S., Antonov, J.I., Boyer, T.P., Locarnini, R.A., Garcia, H.E., and Mishonov, A.V. 2009. Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. Geophys. Res. Lett. 36(7)

Lombard, A.T., Cowling, R.M., Pressey, R.L., and Rebelo, A.G. 2003. Effectiveness of land classes as surrogates for species in conservation planning for the Cape Floristic Region. Biol. Cons. 112(1-2): 45-62

Lund, K., and Wilbur, A.R. 2007. Habitat Classification Feasibility Study for Coastal and Marine Environments in Massachusetts. Massachusetts Office of Coastal Zone Management, Boston, MA.

Lundblad, E.R., Wright, D.J., Miller, J., Larkin, E.M., Rinehart, R., Naar, D.F., Donahue, B.T., Anderson, S.M., and Battista, T. 2006. A Benthic Terrain Classification Scheme for American Samoa. Mar. Geod. 29(2): 89-111

Lyne, V., Fuller, M., Last, P., Butler, A., Martin, M., and Scott, R. 2006. Ecosystem characterisation of Australia's North West Shelf. Technical Report 12. Government of Australia.

Madden, C.J., and Grossman, D.H. 2004. A framework for a coastal/marine ecological classification standard. NatureServe, Arlington, Virginia.

Madden, C.J., Grossman, D.H., and Goodin, K.L. 2005. Coastal and Marine Systems of North America: framework for an ecological classification standard: version II. NatureServe, Arlington, Virginia.

Manson, M.M. 2009. Small Scale Delineation of Northeast Pacific Ocean Undersea Features Using Benthic Position Index. Can. Man. Rep. Fish. Aquat. Sci. 2864

Marchant, R. 1990. Robustness of classification and ordination techniques applied to macroinvertebrate communities from the La Trobe River, Victoria. Australian. J. Mar. Freshwater Res. 41: 493–504

Marchant, R., Hirst, A., Norris, R.H., Butcher, R., Mixzeling, L., and Tiller, D. 1997. Classification and ordination of macroinvertebrate assemblages from running waters in Victoria, Australia. J. N. Am. Bentho. Soc. 16: 664–681

McCune, B., and Grace, J. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon.

McGarigal, K., Cushman, S., and Stafford, S. 2000. Multivariate Statistics for Wildlife and Ecology Research. Springer-Verlag, New York, NY, USA.

McMahon, G., Wiken, E.B., and Gauthier, D.A. 2004. Toward a scientifically rigorous basis for developing mapped ecological regions. Environ. Manage. 34: S111-S124

Milligan, G.W., and Cooper, M.C. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50(2): 159-179

Ministry of Sustainable Resource Management (MSRM) Decision Support Services Branch. 2002. British Columbia Marine Ecological Classification Marine Ecosections and Ecounits Version 2.0. Prepared for the Coastal Task Force Resources Information Standards Committee, Province of British Columbia.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., and Kassem, K.R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Biosci. 51: 933-938

Orr, J.C., Fabry, V.J., Aumont, O., Bopp, L., Doney, S.C., Feely, R.A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R.M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R.G., Plattner, G.K., Rodgers, K.B., Sabine, C.L., Sarmiento, J.L., Schlitzer, R., Slater, R.D., Totterdell, I.J., Weirig, M.F., Yamanaka, Y., and Yool, A. 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437(7059): 681-686

Pitcher, C.R., Lawton, P., Ellis, N., Smith, S.J., Incze, L.S., Wei, C.L., Greenlaw, M.E., Wolff, N.H., Sameoto, J.A., Snelgrove, P.V., and Cadotte, M. 2012. Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. J. Appl. Ecol. 49(3): 670-679

Podani, J., and Csányi, B. 2010. Detecting indicator species: Some extensions of the IndVal measure. Ecol. Ind. 10(6): 1119-1124

Poloczanska, E.S., Brown, C.J., Sydeman, W.J., Kiessling, W., Schoeman, D.S., Moore, P.J., Brander, K., Bruno, J.F., Buckley, L.B., Burrows, M.T., Duarte, C.M., Halpern, B.S., Holding, J., Kappel, C.V., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Schwing, F., Thompson, S.A., and Richardson, A.J. 2013. Global imprint of climate change on marine life. Nat. Clim. Chang. 3(10): 919-925

Poos, M.S., and Jackson, D.A. 2012. Addressing the removal of rare species in multivariate bioassessments: The impact of methodological choices. Ecol. Ind. 18: 82-90

Post, A., Wassenberg, V., and Passlow, V. 2006. Physical surrogates for macrogaunal distributions and abundance in a tropical gulf. Mar. Freshwater Res. 57: 469-483

Prasad, A.M., Iverson, L.R., and Liaw, A. 2006. Newer classification and regression techniques: bagging and random forests for ecological prediction. Ecosystems 9: 181-199

Proches, S. 2005. The world's biogeographical regions: cluster analyses based on bat distributions. J. Biogeog. 32: 607-614

R Core Development Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing,, Vienna, Austria.

Robinson, C., Boutillier, J.A., Biffard, D., Gregr, E.J., Finney, J., Therriault, T., Greenlaw, M.E., Barrie, V., Foreman, M., Pena, A., Masson, D., Bodtker, K.M., Head, K., Spencer, J., Bernhardt, J., Smith, J., and Short, C. 2015. Key elements in the development of a hierarchical marine ecological classification system to support ecosystem approaches to management in Pacific Canada. DFO Can. Sci. Advis. Sec. Res. Doc. 2015/028. viii +58 p.

Robinson, C.L.K., and Royle, K. 2008. An Analysis of Habitat Representation of the Proposed NMCA within the Strait of Georgia Marine Region. Resource Conservation Report. Parks Canada, Vancouver, BC.

Rodrigues, A.S.L., and Brooks, T.M. 2007. Shortcuts for Biodiversity Conservation Planning: The Effectiveness of Surrogates. Annu. Rev. Ecol. Evol. Syst. 38(1): 713-737

Roff, J.C., and Taylor, M.E. 2000. National frameworks for marine conservation — a hierarchical geophysical approach. Aquat. Cons. Mar. Fresh. Ecosys. 10: 209-223

Roff, J.C., Taylor, M.E., and Laughren, J. 2003. Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities. Aquat. Cons. Mar. Fresh. Ecosys. 13(1): 77-90

Roff, J.C., and Zacharias, M.A. 2011. Marine Conservation Ecology. Earthscan, London, UK.

Seo, C., Thorne, J.H., Hannah, L., and Thuiller, W. 2009. Scale effects in species distribution models: implications for conservation planning under climate change. Biol. Lett. 5(1): 39-43

Sheppard, V. 2013. International approaches to characterizing marine seascapes to achieve representativity in MPA network design. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/048. v + 20 p.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M.A.X., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., and Robertson, J. 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. Biosci. 57(7): 573-583

Stevens, T., and Connolly, R.M. 2004. Testing the utility of abiotic surrogates for marine habitat mapping at scales relevant to management. Biol. Cons. 119(3): 351-362

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. 2008. Conditional variable importance for random forests. BMC Bioinform. 9: 307

Sutcliffe, P.R., Klein, C.J., Pitcher, C.R., and Possingham, H.P. 2015. The effectiveness of marine reserve systems constructed using different surrogates of biodiversity. Cons. Biol. 29(3): 657-667

Tulloch, V.J.D., Tulloch, A.I.T., Visconti, P., Halpern, B.S., Watson, J.E.M., Evans, M.C., Auerbach, N.A., Barnes, M., Beger, M., Chadès, I., Giakoumi, S., McDonald-Madden, E., Murray, N.J., Ringma, J., and Possingham, H.P. 2015. Why do we map threats? Linking threat mapping with actions to make better conservation decisions. Front. Ecol. Environ. 13(2): 91-99

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., and De Clerck, O. 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modeling. Global. Ecol. (Accessed 15 March 2016).

UNESCO. 2009. Global Open Oceans and Deep Seabed (GOODS) - Biogeographic Classification. IOC Technical Series, Paris.

Valentine, P.C., Todd, B.J., and Kostylev, V.E. 2005. Classification of Marine Sublittoral Habitats, with Application to the Northeastern North America Region. Am. Fish. Soc. Symp. 41: 183-200

Verfaillie, E., Foster-Smith, B., and Van Lancker, V. 2006. Analysis of bathymetrical derived features on the Belgian Continental Shelf as a support for marine habitat mapping. *In* VLIZ Young Scientists' Day, Brugge, Belgium 31 March 2006: book of abstracts. *Edited by* K. Mees and J. Seys. VLIZ Special Publication. p. 66.

Weiss, A.D. 2001. Topographic Positions and Landforms Analysis (Conference Poster). ESRI International User Conference, San Diego, CA. July 9-13.

Wenger, S.J., Som, N.A., Dauwalter, D.C., Isaak, D.J., Neville, H.M., Luce, C.H., Dunham, J.B., Young, M.K., Fausch, K.D., and Rieman, B.E. 2013. Probabilistic accounting of uncertainty in forecasts of species distributions under climate change. Glob. Chang. Biol. 19(11): 3343-3354

Wiens, J.A. 1989. Spatial Scaling in Ecology. Funct. Ecol. 3: 385-397

Williams, P.H., de Klerk, H.M., and Crowe, T.M. 1999. Interpreting biogeographical boundaries among Afrotropical birds: spatial patterns in richness gradients and species replacement. J. Biogeog. 26: 459-474

Włodarska-Kowalczuk, M., and Kędra, M. 2007. Surrogacy in natural patterns of benthic distribution and diversity: selected taxa versus lower taxonomic resolution. Mar. Ecol. Prog. Ser. 351: 53-63

Wright, D.J., Lundblad, E.R., Larkin, E.M., Rinehart, R.W., Murphy, J., Cary-Kothera, L., and Draganov, K. 2005. ArcGIS Benthic Terrain Modeler. Oregon State University, Davey Jones Locker Seafloor Mapping/Marine GIS Laboratory and NOAA Coastal Services Center, Corvallis, Oregon.

Zacharias, M.A., Howes, D.E., Harper, J.R., and Wainwright, P. 1998. The British Columbia marine ecosystem classification: Rationale, development, and verification. Coast. Manage. 26(2): 105-124

Zacharias, M.A., and Roff, J.C. 2001. Explanations of patterns of intertidal diversity at regional scales. J. Biogeog. 28: 471-483

APPENDIX A. BIOLOGICAL RELEVANCE OF BCMEC ECOSECTIONS

METHODS

To determine if the BCMEC Ecosections (Figure A 1) represented biological diversity patterns in the study area, we assigned each site (4 km grid cell within the Northern Shelf and Southern Shelf Bioregions) to the Ecosection in which its centre point fell. Of the 12 marine Ecosections, the five largest (281 – 1372 sites each in our study area; Continental Slope, Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf), were included in the analysis. The remaining seven Ecosections fell outside of or had limited overlap with our study area (≤ 35 sites each) and were not considered further (data not shown).

We used a permutational analysis of variance (PERMANOVA, Anderson 2001) to test whether the species composition was significantly different among groups (Ecosections) and a test of the homogeneity of multivariate dispersions among groups (PERMDISP, Anderson 2006) to help interpret the PERMANOVA results. A significant PERMANOVA result can be due to differences in centroid location among groups (i.e., differences in species composition), differences in spread (variance), or a combination of the two (Anderson and Walsh 2013). PERMDISP (Anderson 2006) tests if the average within-group dispersion (measured by the average distance to group centroid) is equal among groups. Balanced PERMANOVA tests are more robust (Anderson and Walsh 2013), so we randomly resampled the number of sites in each Ecosection to the smallest sample size (n=281, Hecate Strait). Each test was run with 999 permutations.

To aid in the interpretation of the results, we examined the data visually, using non-metric multidimensional scaling (nMDS). nMDS is an iterative search for a ranking and placement of *n* entities on *k* dimensions (axes) that minimizes the stress of the k-dimensional configurations, where stress is a measure of departure from monotonicity in the relationship between the distance in the original matrix and distance in the reduced k-dimensional ordination space (McCune and Grace 2002). The nMDS plot provides a visualization of the differences in species composition among groups. PERMANOVA, PERMDISP, and nMDS were run in R using the *adonis*, *betadisper,* and *metaMDS* functions in the 'vegan' package (Oksanen 2014).

RESULTS

The PERMANOVA results revealed significant differences in species composition among Ecosections (F = 166.83, df = 4, $p < 0.001$; Table A 1). However, the PERMDISP test rejected the null hypothesis of homogeneity of multivariate dispersion among all groups (F = 5.8, df = 4, $p < 0.001$; Table A 2) indicating that the significant PERMANOVA result could be driven by differences in multivariate spread in the data within groups. The effect size (R^2) shows that only 32% of the variation is explained by Ecosections leaving 68% of the variation explained within groups.

We used an nMDS plot to examine the similarity of Ecosections in multidimensional space (Figure A 2a). The plot, showing the 95% ellipses, indicates high overlap between three Ecosections (Dixon Entrance, Queen Charlotte Sound, and Vancouver Island Shelf), with only the Continental Slope Ecosection, and to a lesser degree, Hecate Strait, showing any distinction from the others. To determine if the continental slope was driving the significant results of the PERMANOVA, we removed it and reran the analysis. With continental slope removed, the PERMANOVA showed a significant result (F = 80.3, df = 3, $p < 0.001$; Table A 3), but the amount of variation among groups decreased from 32% to 18%.

Our results showed that the physiographic classification reflects significant compositional turnover in benthic species, however, further analyses showed there are few species with strong associations to most of the Ecosections. Also, the nMDS analysis showed that there is high compositional overlap among Ecosections and only the Continental Slope Ecosection displays a visually distinct assemblage. This suggests that if this classification was used in MPA planning to fulfil the representativity criterion, the species associated with Slope habitat would be captured in the network, but because changes in benthic diversity do not appear to be structured by the other Ecosection boundaries, distinct assemblages over the rest of the continental shelf would not be adequately considered in the planning process.

To compare the biological representativity of BCMEC to the PMECS Level 4b biophysical units, we ran PERMANOVA and nMDS analyses as above. Using a balanced design (all biophysical units randomly resampled to $n = 120$), we found significant differences in species assemblages, among biophysical units ($F = 221.34$, $df = 4$, $p < 0.0001$; Table A 4). However, the PERMDISP test rejected the null hypothesis of homogeneity of multivariate dispersion among all groups ($F = 8.077$, $df = 4$, $p < 0.001$; Table A 5) indicating that the significant PERMANOVA result could be driven by the significant spread in the data within groups. However, the PERMANOVA results show that 60% of the variation is explained among biophysical units and 40% of the variation explained within groups. This is in contrast to the results of the Ecosections PERMANOVA, where the majority of the variation (68%) was due to variation within Ecosections and only 32% was due to variation among Ecosections.

The nMDS plot (Figure A 2b), shows considerably less overlap in composition among biophysical units than for the Ecosections. Most of the overlap occurs only at the transitions between spatially neighbouring biophysical units with the highest overlap between the Dogfish Bank and Other Banks biophysical units. The Shelf biophysical unit shows overlap with all other units except for the Slope. The Slope biophysical unit, similarly to the Continental Slope Ecosection, is the most distinct assemblage with only a small amount of overlap with the Troughs biophysical unit.

Table A 1. PERMANOVA analysis of variance table for Ecosections analysis, including 5 Ecosections (Continental Slope, Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf).

PERMANOVA Analysis of Variance	Df	Sums Of Sqs	Mean Sq	F.Model	R^2	Pr(>F)
Ecosections	4	97.702	24.4255	166.83	0.32279	0.001
Residuals	1400	204.976	0.1464	0.67721	-	-
Total	1404	302.677	1	-	-	-

Table A 2. PERMDISP analysis of variance table for Ecosections analysis, including 5 Ecosections (Continental Slope, Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf).

PERMDISP Analysis of Variance	Df	SumSq	Mean Sq	F-Value	Pr(>F)
Ecosections	4	0.4601	0.115033	5.8241	< 0.001
Residuals	1400	27.6518	0.019751	-	-

Table A 3. PERMANOVA analysis of variance table for Ecosections analysis, including 4 Ecosections (Dixon Entrance, Hecate Strait, Queen Charlotte Sound, and Vancouver Island Shelf).

PERMANOVA Analysis of Variance	Df	Sums Of Sqs	Mean Sq	F.Model	R^2	Pr(>F)
Ecosections	3	34.342	11.4474	80.306	0.17703	0.001
Residuals	1120	159.652	0.1425	-	0.82297	-
Total	1123	193.994	-	-	1	-

Table A 4. PERMANOVA analysis of variance table for Level 4b biophysical unit analysis.

PERMANOVA Analysis of Variance	Df	Sums Of Sqs	MeanSqs	F.Model	R^2	Pr(>F)
Biophysical units	4	98.842	24.7105	221.34	0.59807	0.001
Residuals	595	66.428	0.1116	-	0.40193	-
Total	599	165.27	-	-	1	-

Table A 5. PERMDISP analysis of variance table for Level 4b biophysical unit analysis.

PERMDISP Analysis of Variance	Df	SumSq	MeanSq	F Value	Pr(>F)
Biophysical units	4	0.5757	0.143925	8.0777	< 0.001
Residuals	595	10.6014	0.017817	-	-

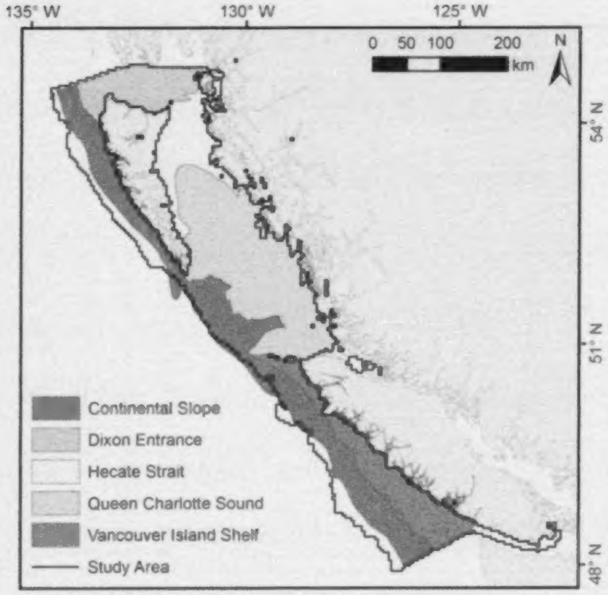

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Continental Slope
Dixon Entrance
Hecate Strait
Queen Charlotte Sound
Vancouver Island Shelf
Study Area

Figure A 1. Subset of British Columbia Marine Ecological Classification (BCMEC) Ecosections analyzed in this paper. Seven additional marine Ecosections exist in BCMEC (Zacharias et al. 1998) but are not considered here.

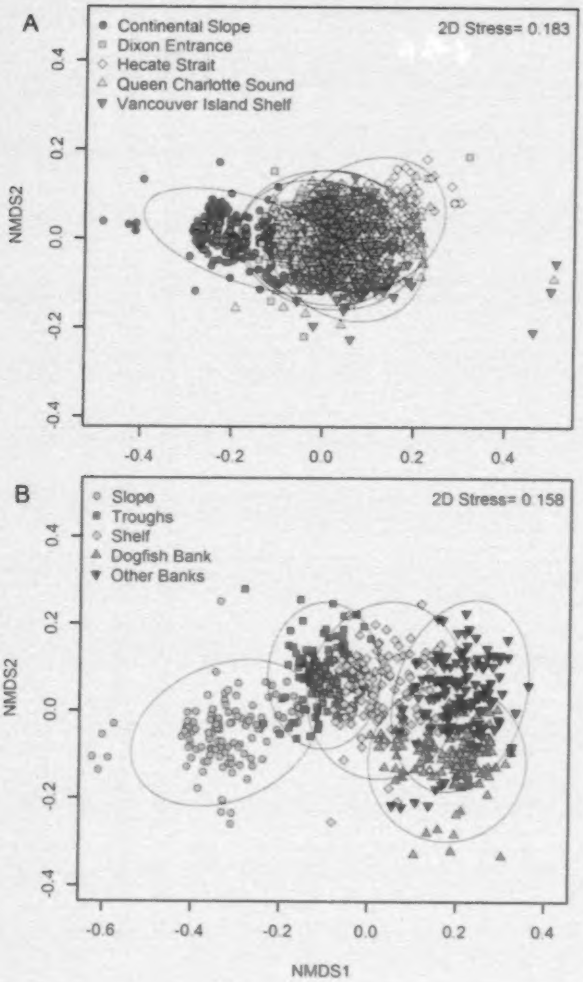


Figure A 2. Non-metric multidimensional scaling plot of A) BCMEC Ecosections and B) Level 4b biophysical units, with 95% confidence ellipses.

APPENDIX B. DETAILS ON PRIMARY DATA COLLECTION AND PREPARATION

SHELLFISH DATA HOLDINGS (SF LOG AND SF BIO)

Table B 1. Description, years, and number of records and taxa from databases within the DFO Shellfish Log (SF Log; commercial fisheries) data holdings. Taxa numbers do not sum because of shared taxa among databases. "In PMECS Geodatabase" refers to data that were processed and name-verified, and contains all taxonomic levels for all phyla and habitats. "In analysis" refers to data that were used in the cluster analyses and which were limited to species-level records of benthic and demersal species within the study area. "Record" refers to a single record of a single species; this is equivalent to the number of lines in the database if each line includes only one species.

		In PMECS Geodatabase		In analysis	
Database	Years extracted	No. records	No. taxa	No. records	No. species
Crab by Trap	2000–14	338,886	4	Excluded – limited number of species	
Geoduck	2000–14	86,465	1	Excluded – limited number of species	
Green Urchin	2001–12	2,585	1	Excluded – limited number of species	
Horse Clam	2000–14	1,982	1	Excluded – limited number of species	
Octopus by Dive	2000–13	4,363	1	Excluded – limited number of species	
Prawn by Trap	2000–14	844,082	3	Excluded – limited number of species	
Red Urchin	2001–12	31,814	1	Excluded – limited number of species	
Scallop by Trawl	2001–14	1569	36	Excluded – limited spatial extent	
Sea Cucumber	2001–13	15,272	1	Excluded – limited number of species	
Shrimp by Trawl	2000–14	247,834	20	Excluded – patchy spatial extent	
Squid	2000–05	58	1	Excluded – limited number of species	
Tanner Crab	2000–03	39,120	54	Excluded – contains fewer species than the Tanner Crab research (see Table B 2), which we considered more comprehensive	
Total for SF Log	-	**1,614,010**	**104**	-	-

Table B 2. Description, years, and number of georeferenced records and taxa from databases within the DFO Bio (SF Bio; research) data holdings. Taxa numbers do not sum because of shared taxa among databases. "In PMECS geodatabase" refers to data that were processed and name-verified, and contains all taxonomic levels for all phyla and habitats. "In analysis" refers to data that were used in the cluster analysis and which were limited to species-level records of benthic and demersal species within the study area. "Record" refers to a single record of a single species; this is equivalent to the number of lines in the database if each line includes only one species.

		In PMECS geodatabase		In analysis	
Database	Years	No. records	No. taxa	No. records	No. species
Crab by Trap	2000–14	794,631	330	133,007	26
Tanner Crab	2000–06	10,797	499	4,296	111
Intertidal Clam	2000–13	103,379	369	Excluded – intertidal/nearshore only	
Geoduck	2000–13	8,396	4	Excluded – limited number of species	
Green Urchin	2000–14	15,991	30	Excluded – limited spatial extent	
Prawn by Trap	2000–14	31,826	4	Excluded – nearshore only, and limited number of species	
Red Urchin	2000–14	51,633	6	Excluded – nearshore only, and limited number of species	
Scallop Trawl	2000–11	733	57	Excluded – limited spatial extent	
Sea Cucumber	2000–14	169,300	44	Excluded – nearshore only	
Sea Cucumber database – geoduck records	2010–14	1,463	1	Excluded – limited number of species	
Shrimp	2000–14	153,988	554	Excluded – patchy spatial extent	
Total for SF Bio	-	**1,342,137**	**1,214**	**137,303**	**137**

Crab by Trap

Description: The crab trap fishery for Dungeness Crab (*Metacarcinus magister*) takes place in most coastal areas of BC. A number of research datasets exist related to crab research and fishery monitoring, including data from fishery-dependent and fishery-independent sampling, DFO research, and fisheries observers. DFO sets research trap lines using commercial style circular metal Dungeness traps (90 cm diameter, 26 cm high) with closed escape ports; traps are baited with herring and soaked overnight for approximately 24 hours (Dunham et al. 2011). Data from an established soft-shell monitoring program in Area A (Queen Charlotte Sound, Hecate Strait, Haida Gwaii) is included in the crab trap research database to provide information on moult timing (DFO Crab by Trap IFMP 2015).

Data Locations: All data from the crab by trap database were considered for use in the cluster analysis. The data used for the analysis mostly came from Dogfish Bank, with some points on Goose Bank (Queen Charlotte Sound) and in the Strait of Juan de Fuca.

Reason for use: The crab trap research data includes a large number of points (133,007 of 234,842 records used in final cluster analysis, 56%). Of the 62 sites that this dataset contributes to, most are around Dogfish Bank. In addition, the crab trap research dataset populates 17 unique sites (sites where no other biological data was available), 11 of which are on Dogfish Bank. Many of the programs that contribute to the crab trap research database collect data on co-occurring species (i.e., not just Dungeness Crab), and so satisfy our criteria of "multispecies surveys". The exception is data from the softshell monitoring program, which we used in our analysis despite it containing only single-species records of Dungeness Crab, to improve spatial coverage of Dogfish Bank.

Tanner Crab

Description: Research surveys for Tanner Crab (*Chionoecetes tanneri*) were carried out between 1999-2006 as part of the experimental Tanner Crab commercial fishery that ran between 1999-2003 (Gillespie et al. 2004). The research trawl surveys were carried out using a Campelen 1800 shrimp trawl with rockhopper footgear, and supplementary trap surveys were carried out between 1999-2001 with top-loading conical traps with 40 mm mesh. Tanner crab research survey data were gathered in PFMAs 23, 101, 102, 123-127, 130, and 142. Incidental species were also recorded.

Data Locations: Data from the tanner crab research surveys were used in the cluster analysis. This data mostly came from the continental slope of BC, with some points in Barkley Sound and Dixon Entrance.

Reason for use: Although the Tanner Crab dataset makes up a small portion of our records (only 4296 of 234,842 records used in final cluster analysis, < 2%), the continental slope area had otherwise poor data coverage. The Tanner Crab dataset contributes data to 155 sites, including 82 unique sites.

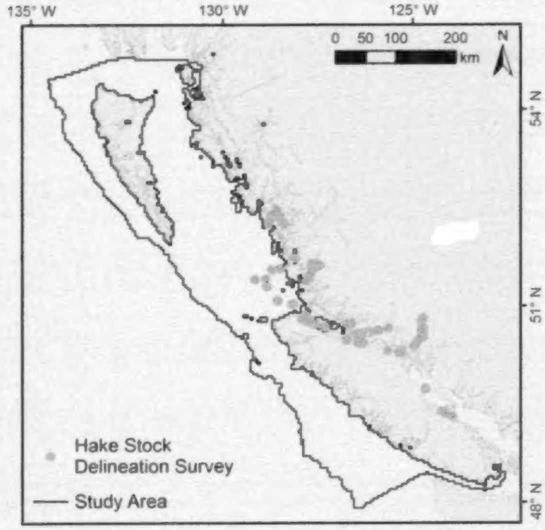
GROUNDFISH DATABASES (GFBIO)

Table B 3. Description of each DFO groundfish survey data source including survey name and data details. Data shown are for all georeferenced records with valid taxonomic names, including all taxa (identified to species and identified to higher taxa) in all habitats. "In working paper analysis" refers to data that were used in the cluster analysis and which were limited to species-level records of benthic and demersal species within the study area. "Record" refers to a single record of a single species; this is equivalent to the number of lines in the database if each line includes only one species.

		In PMECS geodatabase		In analysis	
Survey Name	Years	No. records	No. taxa	No. records	No. spp.
Georgia Strait Ecosystem Research Initiative	2011	236	38	-	-
Hake Stock Delineation	2007–08	525	60	42	12
Hecate Strait Multispecies Trawl	2003	2,068	144	1648	75
Hecate Strait Pacific Cod Monitoring Trawl	2003–04	484	61	417	38
Hecate Strait Synoptic Trawl	2005, 07, 09, 11, 13	19,272	427	15,072	141
Inshore Rockfish Longline	2003–2013	3,452	94	153	21
International Pacific Halibut Commission Longline	2003–2012	14,078	124	10,762	58
Joint Can/US Hake Acoustic	2003, 05, 07, 09, 11–13	3604	274	576	71
Pacific Halibut Management Association Longline	2006–2012	14,087	139	8,485	65
Queen Charlotte Sound Synoptic Trawl	2003–05, 07, 09, 11, 13	36,412	507	28,573	151
Sablefish Research and Assessment	2003–13	10,259	218	7,544	88
Strait of Georgia Synoptic Trawl	2012	1,259	137	-	-
West Coast Queen Charlotte Island Synoptic Trawl	2006–08, 10, 12	12,990	410	9,776	134
West Coast Vancouver Island Synoptic Trawl	2004, 06, 08, 10, 12	16,183	365	13,462	143
West Coast Vancouver Island Thornyhead Trawl	2003	1,361	93	787	37
Yelloweye Rockfish Charter Longline	2003	451	32	242	19
Total GFBIO	-	**136,721**	**868**	**97,539**	**169**

Hake Stock Delineation

This survey took place to assess a spatial shift in hake populations. Midwater trawl sets supplemented acoustic surveys at depths of 82-800 m in the Strait of Georgia and Queen Charlotte Sound (King et al. 2012).

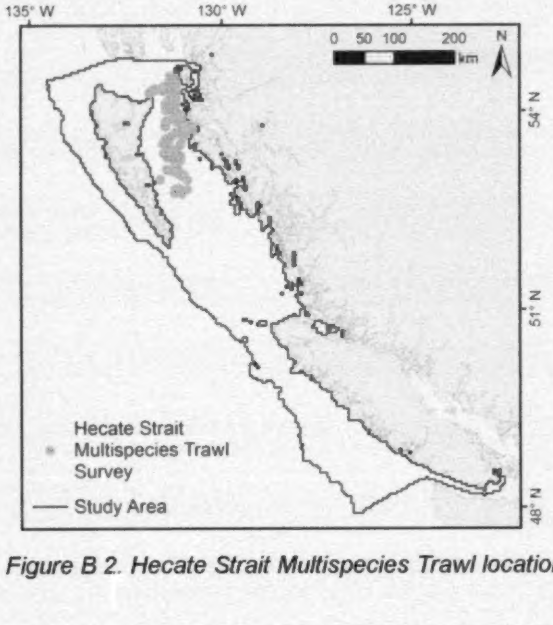


135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Hake Stock Delineation Survey
Study Area

Figure B 1. Hake Stock Delineation Survey locations.

Hecate Strait Multispecies Trawl Survey

The Hecate Strait multispecies trawl survey included Hecate Strait and part of Dixon Entrance, at depths of 18–146 m over a sampling grid of 10 x 10 nautical miles (Choromanski et al. 2005). A Yankee 36 net with a 3.5 inch mesh was used with 30 minute tows. All catch was sorted by species and weighed, with biological sampling for target species (Choromanski et al. 2005).



135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Hecate Strait Multispecies Trawl Survey
Study Area

Figure B 2. Hecate Strait Multispecies Trawl locations.

Hecate Strait Pacific Cod Monitoring Trawl

This was a stratified random bottom trawl survey, using a 0.01 degree latitude x 0.01 degree longitude grid. The gear was an Atlantic Western IIA bottom trawl with 4.5 inch mesh, with target tow duration between 20 and 30 minutes. Catch was sorted to lowest possible taxonomic level and weighed, and additional biological data were collected for Pacific cod (Sinclair and Workman 2002).

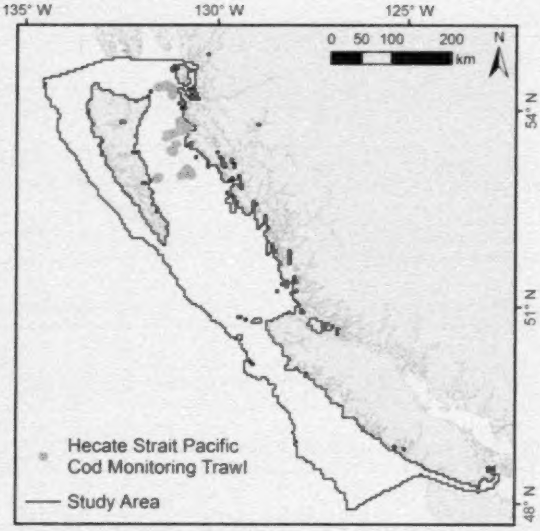


Figure B 3. Hecate Strait Pacific Cod Monitoring Trawl locations.

Inshore Rockfish Longline Survey

A depth-stratified random design using 2 km x 2 km grid cells was used over depths of 41–100 m. Flat, mud, and sandy bottom were avoided. Snap-type longline gear with two skates of groundline was used, with squid-baited circle hooks and perlon gangions spaced 12 feet apart (Lochead and Yamanaka 2007). Catch was recorded hook-by-hook to species, with extra biological data collected for target species (Lochead and Yamanaka 2007).

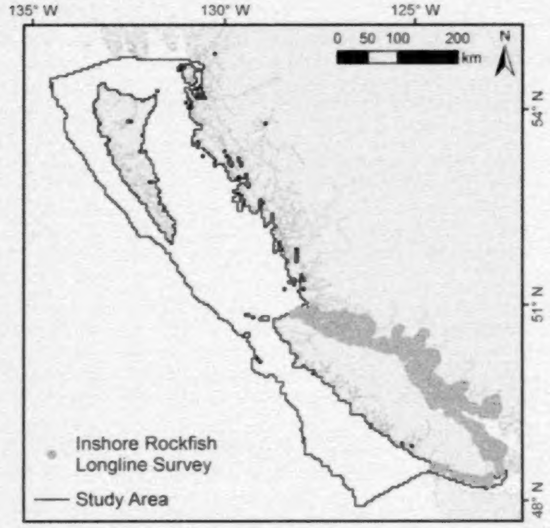

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Inshore Rockfish Longline Survey
Study Area

Figure B 4. Inshore Rockfish Longline Survey locations.

International Pacific Halibut Commission (IPHC) Longline Survey

The IPHC surveys cover nearshore and offshore waters from Oregon to the Bering Sea and include170 non-random survey stations in BC (White et al. 2010, Flemming et al. 2012). The standard IPHC gear includes fixed-hook, 1,800-foot skates (5-8 skates per set depending on year, Flemming et al. 2012) with 100 circle hooks baited with Chum Salmon and placed 18 feet apart (White et al. 2010). Starting in 2003, an additional technician was employed on the IPHC surveys to collect multispecies records (on a "hook by hook" basis), as well as biological data on rockfish and Sablefish (Yamanaka et al. 2004b, Flemming et al. 2012).

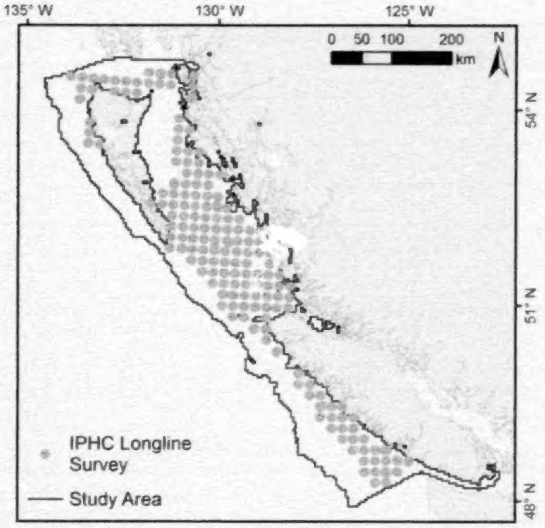

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
IPHC Longline Survey
Study Area

Figure B 5. International Pacific Halibut Commission (IPHC) Longline Survey locations.

Joint Canada/US Hake Acoustic Survey

Trawls were carried out to complement acoustic surveys and were not systematic in design. Pelagic trawls were performed with a Polish rope trawl and near-bottom trawls with a Yankee 36 research trawl modified with roller gear (Fleischer et al. 2005), for an average of 21 minutes. The near-bottom trawls attempted to avoid hard bottoms, flatfish, and other benthic fish. Catch was sorted and weighed, and biological data collected for Pacific Hake (Fleischer et al. 2005).

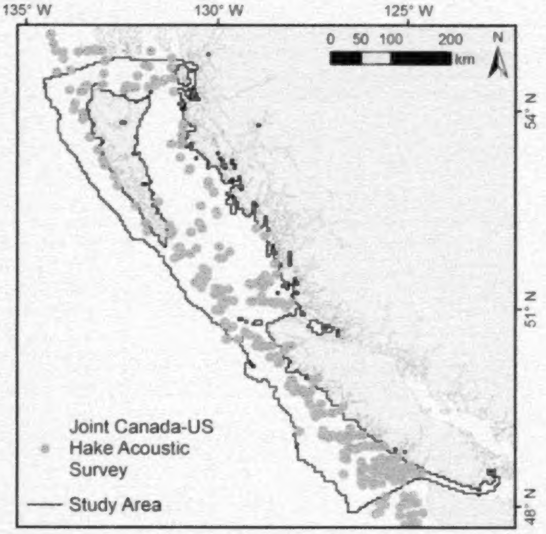

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Joint Canada-US Hake Acoustic Survey
Study Area

Figure B 6. Joint Canada/US Hake Acoustic Survey locations.

Pacific Halibut Management Association (PHMA) Longline Survey

This is a depth stratified, random designed longline survey targeting hard bottomed areas and using the same grid as for the synoptic trawl surveys (DFO Groundfish IFMP 2011). This survey alternates annually between northern and southern areas of coastal BC. There is a target of 200 fishing sets selected in each year and three commercial vessels are chartered to fish in one of three areas within the northern or southern portion each year (K. Rutherford, DFO, Nanaimo, B.C., personal communication, 2015).

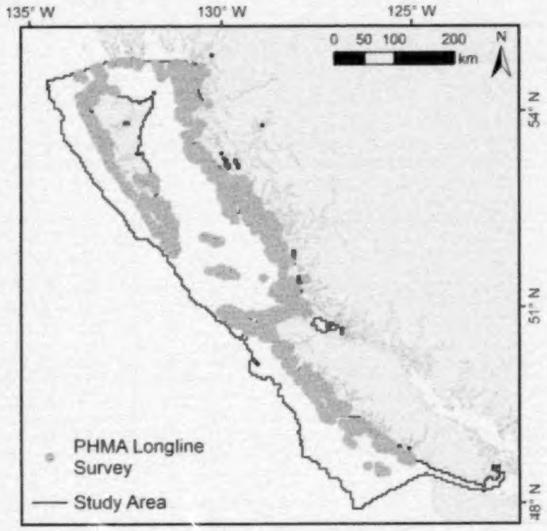


Figure B 7. Pacific Halibut Management Association (PHMA) Longline Survey locations.

Sablefish Research and Assessment Survey

Standardized sets occur at offshore (up to 2010) and inlet index locations, as well as at locations within stratified random locations in depth and spatial strata. Depths between 91-1389 m were sampled at index sites, whereas the stratified random samples were taken between 183-1372 m (Wyeth et al. 2007). Sampling units of 2 km by 2 km grid cells were used. Sets were made using longline trap gear typical of that used in the commercial sablefish trap industry; this involved a groundline running along the sea floor, anchored at each end, with 25 squid-baited traps set along 46 m intervals. The weight or count of all species in each trap is recorded, as well as biological data on the Sablefish (Wyeth et al. 2007). Changes have been made to this survey design as of the 2015 survey.

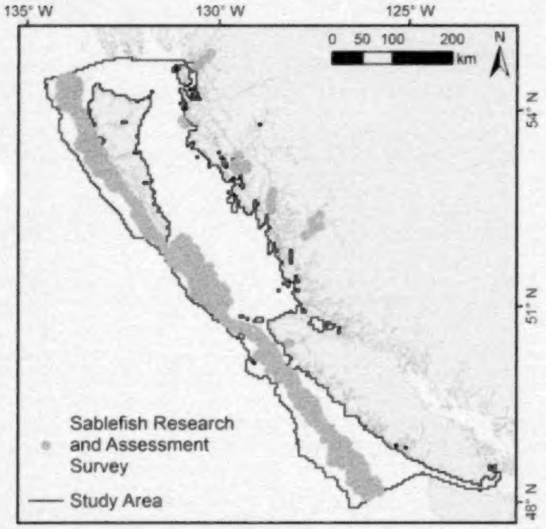


Figure B 8. Sablefish Research and Assessment Survey locations.

Synoptic Surveys (Hecate Strait, Queen Charlotte Sound, West Coast Vancouver Island, West Coast Queen Charlotte Island)

Sixty-two percent of the records in the GFBio dataset came from the Pacific groundfish synoptic surveys. Together, these surveys comprehensively cover the continental shelf of BC (Olsen et al. 2009b). The surveys do not include inlets, enclosed waters, sensitive habitats (e.g., Hecate Strait Glass Sponge reefs, Learmonth Bank Red Tree corals, Rockfish Conservation Areas), or areas that are not trawlable (Workman et al. 2008a, Olsen et al. 2009b).

The surveys follow a stratified random design, with each of the regions being split into depth and area strata. A proportional random sample of 4 km^2 (2km x 2km) grid cells within each stratum are selected to be surveyed (Olsen et al. 2009b). Depths between 50 and 500 m are divided into 4 depth strata (Stanley et al. 2007, Olsen et al. 2009a) in all regions except the west coast of Haida Gwaii (West Coast Queen Charlotte Island), where the depth range is 180–1300 m (Workman et al. 2008c). Tow length is 20 minutes, except for deep Hecate tows (>500 m) which are 40 minutes (Workman et al. 2008a, Olsen et al. 2009b) and for the deepest tows off of west coast of Haida Gwaii (>800 m), which were 40 minutes until 2008, after which time they were 20 minutes. The standard gear is an Atlantic Western IIA box trawl with 5 inch mesh.

Catch is sorted by lowest possible taxonomic group and weighed or counted; subsamples are sometimes taken, and most fishes are identified to species, while invertebrates are often recorded to higher taxa. Biological data (size, sex, age) are also collected for species of interest (Workman et al. 2008c, Olsen et al. 2009b).

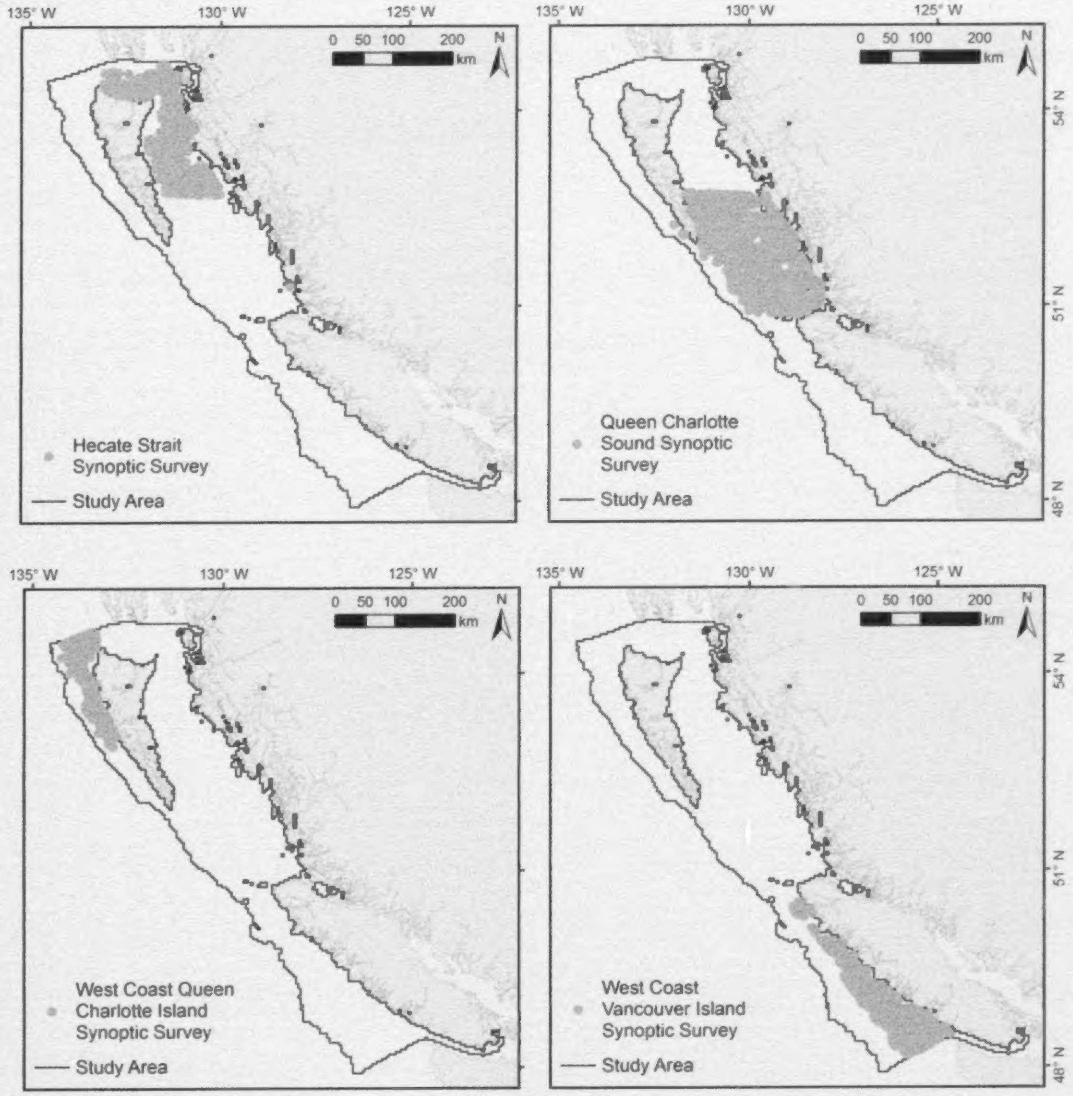

135° W
130° W
125° W
0 50 100 200 km
N
54° N
51° N
48° N
Hecate Strait Synoptic Survey
Queen Charlotte Sound Synoptic Survey
West Coast Queen Charlotte Island Synoptic Survey
West Coast Vancouver Island Synoptic Survey
Study Area

Figure B 9. Synoptic Survey locations (Clockwise from upper left: Hecate Strait, Queen Charlotte Sound, West Coast Vancouver Island, West Coast Queen Charlotte Island)

West Coast Vancouver Island Thornyhead Trawl Survey

Randomly stratified stations (500 m x 500 m grids, depths of 500–1600 m) were selected to be trawled using a standard Atlantic Western II box trawl with a 4.5 inch mesh, with a standard tow length of 30 minutes (Krishka et al. 2005). All catch was sorted by taxonomic group and weighed or counted, with extra data collected for target species (Krishka et al. 2005).

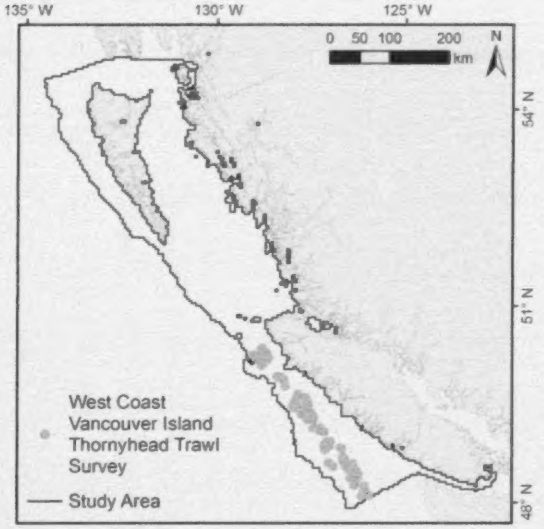

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
West Coast Vancouver Island Thornyhead Trawl Survey
Study Area

Figure B 10. West Coast Vancouver Island Thornyhead Trawl Survey locations.

Yelloweye Rockfish Charter Longline Survey

Fishing for Yelloweye occurred at four index sites on the lower west coast of Haida Gwaii and the upper west coast of Vancouver Island, at depths of 40-200 m. Longline gear with 8-foot spaced squid-baited hooks was used (Yamanaka et al. 2004a).

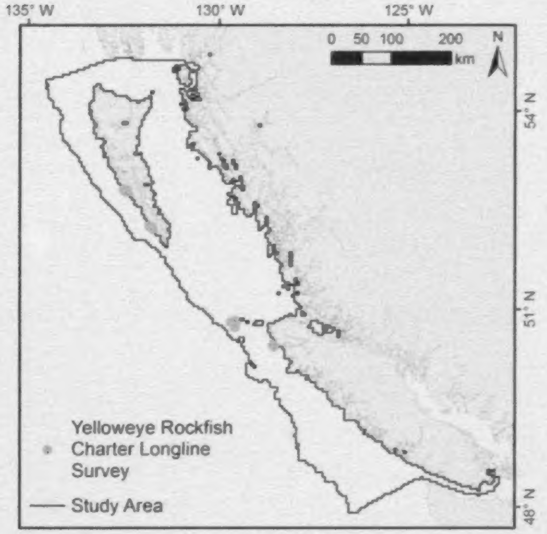

135° W
130° W
125° W
0 50 100 200
km
N
54° N
51° N
48° N
Yelloweye Rockfish Charter Longline Survey
Study Area

Figure B 11. Yelloweye Rockfish Charter Longline Survey locations.

APPENDIX C. CORRELATION ANALYSIS

Of the 59 environmental variables prepared for random forest analysis, many were highly correlated (-0.82 to 0.99, Figure C 1). To reduce the number of explanatory variables in the random forest we dropped variables based on high correlations, limited extent, and our predicted ecological relevance of each variable. We retained 14 environmental variables (Figure C 2) with correlation values ranging from -0.68 to 0.62.

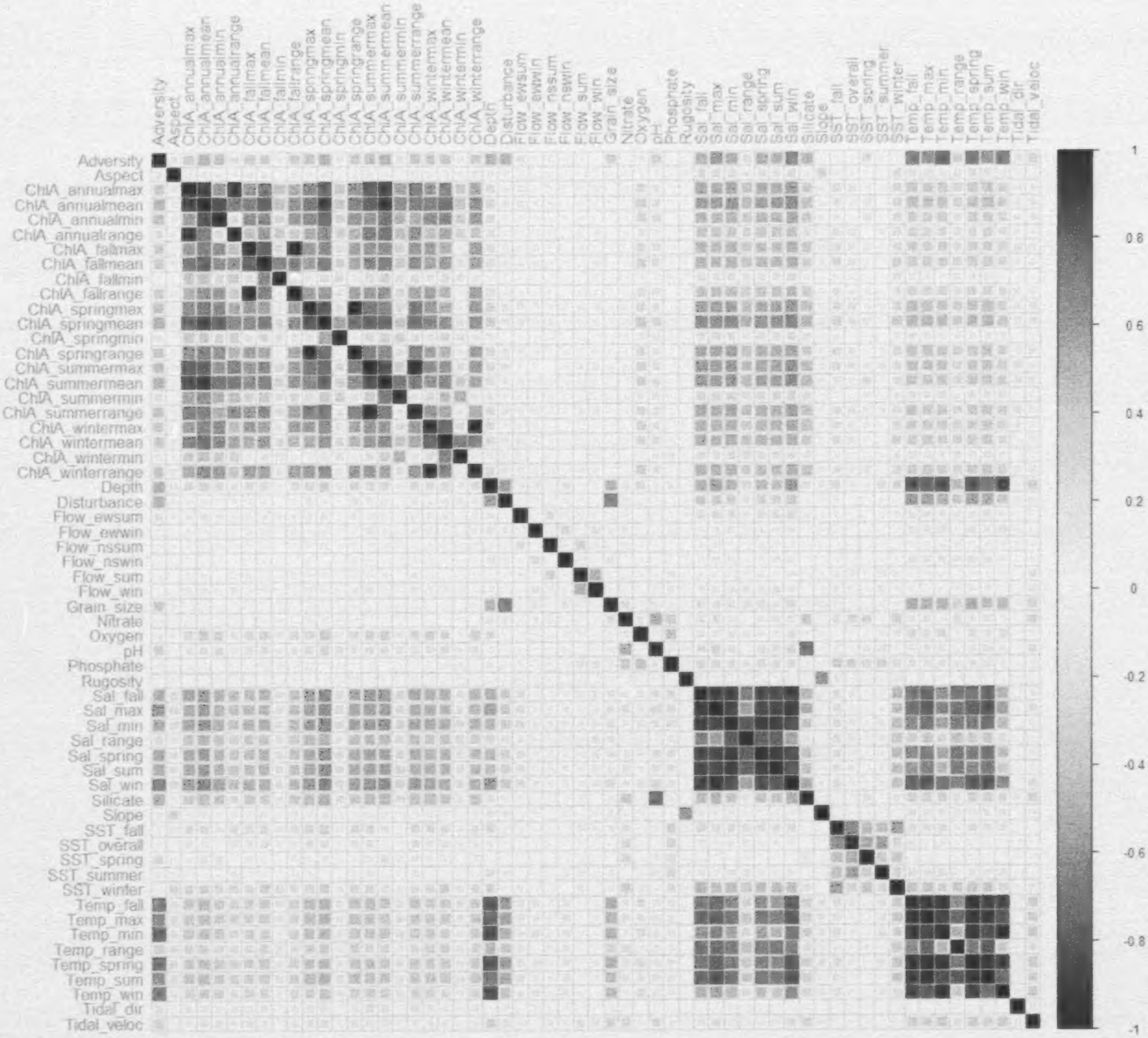


Figure C 1. Correlation plot of 59 environmental variables at 4-km resolution.

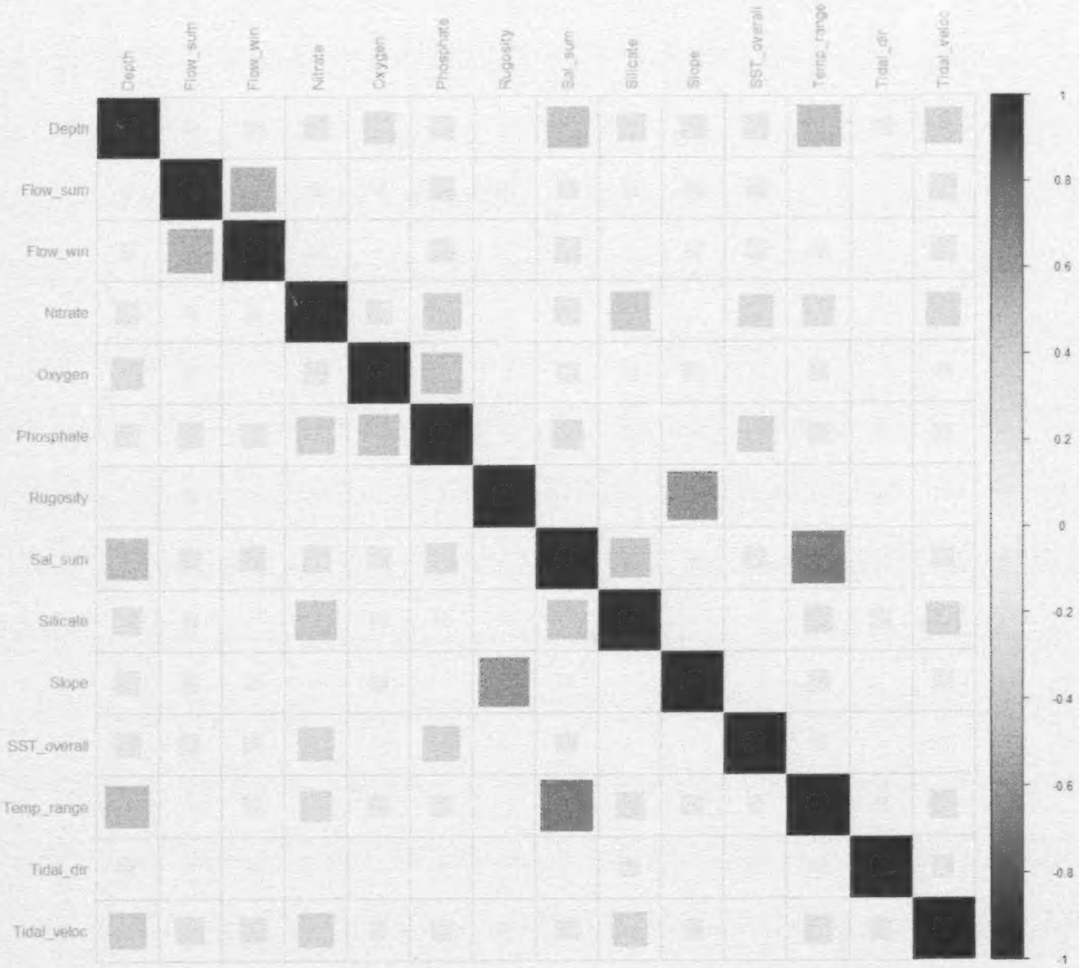


Figure C 2. Correlation plot of final 14 environmental variables that were used in the random forest analyses.

APPENDIX D. SPECIES INFORMATION

Table D 1. Species included in cluster analysis for Level 4a and 4b biophysical unit identification. These species belonged to an acceptable taxon with an acceptable habitat, were recorded in the Groundfish, Crab, or Tanner Crab research surveys, and were represented in the final analysis. Threshold for inclusion was species found in more than 1% of sites (37 sites). Habitats were obtained from FishBase (2016), SeaLifeBase (2016), and other sources and were accurate from those sources as of December 2014.

Phylum	Species	Habitat	Number of sites present
Decapods (crabs, shrimp; 26 spp.)	*Argis alaskensis*	Benthic	51
	Argis levior	Benthic	40
	Chionoecetes angulatus	Benthic	50
	Chionoecetes tanneri	Benthic	482
	Chorilia longipes	Benthic	87
	Crangon alaskensis	Benthic	114
	Eualus biunguis	Benthic	79
	Eualus macrophthalmus	Benthic	77
	Lithodes couesi	Benthic	272
	Lopholithodes foraminatus	Benthic	42
	Metacarcinus magister	Benthic	192
	Munida quadrispina	Benthic	126
	Neocrangon communis	Benthic	100
	Notostomus japonicus	Benthic	92
	Oregonia gracilis	Benthic	108
	Pandalopsis dispar	Benthic	610
	Pandalus borealis	Benthic	228
	Pandalus danae	Benthic	48
	Pandalus jordani	Benthic	768
	Pandalus platyceros	Benthic	631
	Pandalus tridens	Benthic	137
	Paracrangon echinata	Benthic	97
	Paralomis multispina	Benthic	93
	Pasiphaea pacifica	Benthic	209
	Pasiphaea tarda	Benthic	70
	Romaleon branneri	Benthic	55
Fish (96 spp.)	*Albatrossia pectoralis*	Demersal	474
	Ammodytes hexapterus	Benthopelagic	246
	Anarrhichthys ocellatus	Demersal	80
	Anoplopoma fimbria	Demersal	2167
	Apristurus brunneus	Demersal	109
	Atheresthes stomias	Demersal	2503
	Bathyagonus nigripinnis	Demersal	119
	Bathyagonus pentacanthus	Demersal	258
	Bathyraja abyssicola	Demersal	45

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Bathyraja aleutica*	Demersal	125
	Bathyraja interrupta	Demersal	443
	Bathyraja parmifera	Demersal	46
	Bathyraja trachura	Demersal	128
	Bothrocara brunneum	Demersal	143
	Careproctus melanurus	Demersal	104
	Chitonotus pugetensis	Demersal	141
	Citharichthys sordidus	Demersal	551
	Coryphaenoides acrolepis	Demersal	462
	Coryphaenoides cinereus	Demersal	157
	Coryphaenoides filifer	Demersal	77
	Cymatogaster aggregata	Demersal	115
	Embassichthys bathybius	Demersal	156
	Enophrys bison	Demersal	52
	Entosphenus tridentatus	Demersal	62
	Eopsetta jordani	Demersal	1164
	Eptatretus deani	Demersal	72
	Gadus macrocephalus	Demersal	1720
	Glyptocephalus zachirus	Demersal	1860
	Hemilepidotus hemilepidotus	Demersal	38
	Hemitripterus bolini	Demersal	61
	Hexagrammos decagrammus	Demersal	170
	Hippoglossoides elassodon	Demersal	772
	Hippoglossus stenolepis	Demersal	2076
	Hydrolagus colliei	Demersal	2347
	Icelinus filamentosus	Demersal	313
	Icelinus tenuis	Demersal	64
	Isopsetta isolepis	Demersal	174
	Lepidopsetta bilineata	Demersal	678
	Lumpenus sagitta	Benthopelagic	149
	Lycenchelys crotalinus	Demersal	59
	Lycodes brevipes	Demersal	61
	Lycodes cortezianus	Demersal	190
	Lycodes diapterus	Demersal	276
	Lycodes pacificus	Demersal	578
	Lycodes palearis	Demersal	68
	Lyopsetta exilis	Demersal	1088
	Malacocottus kincaidi	Demersal	244
	Malacocottus zonurus	Demersal	251
	Microgadus proximus	Demersal	249
	Microstomus pacificus	Demersal	1831
	Oncorhynchus keta	Benthopelagic	75

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Oncorhynchus tshawytscha*	Benthopelagic	83
	Ophiodon elongatus	Demersal	1644
	Parophrys vetulus	Demersal	1117
	Platichthys stellatus	Demersal	40
	Pleuronichthys decurrens	Demersal	289
	Podothecus accipenserinus	Demersal	230
	Poroclinus rothrocki	Demersal	53
	Psettichthys melanostictus	Demersal	196
	Psychrolutes paradoxus	Demersal	61
	Radulinus asprellus	Demersal	120
	Raja binoculata	Demersal	613
	Raja rhina	Demersal	1645
	Ronquilus jordani	Demersal	145
	Sebastes aleutianus	Demersal	869
	Sebastes alutus	Demersal	1248
	Sebastes aurora	Demersal	84
	Sebastes babcocki	Demersal	1303
	Sebastes borealis	Demersal	346
	Sebastes brevispinis	Demersal	1437
	Sebastes caurinus	Demersal	139
	Sebastes crameri	Demersal	298
	Sebastes diploproa	Demersal	391
	Sebastes elongatus	Demersal	843
	Sebastes flavidus	Demersal	798
	Sebastes helvomaculatus	Demersal	719
	Sebastes jordani	Demersal	41
	Sebastes maliger	Demersal	672
	Sebastes miniatus	Reef-associated	45
	Sebastes nebulosus	Reef-associated	96
	Sebastes nigrocinctus	Reef-associated	103
	Sebastes paucispinis	Reef-associated	384
	Sebastes pinniger	Demersal	902
	Sebastes proriger	Demersal	664
	Sebastes reedi	Demersal	503
	Sebastes ruberrimus	Reef-associated	935
	Sebastes variegatus	Demersal	162
	Sebastes wilsoni	Demersal	229
	Sebastes zacentrus	Demersal	794
	Sebastolobus alascanus	Demersal	1265
	Sebastolobus altivelis	Demersal	335
	Somniosus pacificus	Benthopelagic	42
	Theragra chalcogramma	Benthopelagic	1206

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Triglops macellus*	Demersal	49
	Xeneretmus latifrons	Demersal	51
	Xeneretmus leiops	Demersal	103
Anthozoans (sea pens, corals; 3 spp.)	*Halipteris willemoesi*	Benthic	287
	Paragorgia arborea	Benthic	47
	Ptilosarcus gurneyi	Benthic	88
Asteroids (sea stars; 22 spp.)	*Ceramaster patagonicus*	Benthic	53
	Cheiraster (Luidiaster) dawsoni	Benthic	140
	Crossaster borealis	Benthic	112
	Crossaster papposus	Benthic	246
	Ctenodiscus crispatus	Benthic	235
	Dermasterias imbricata	Benthic	50
	Heterozonias alternatus	Benthic	66
	Hippasteria californica	Benthic	86
	Hippasteria phrygiana	Benthic	183
	Lophaster furcilliger	Benthic	84
	Luidia foliolata	Benthic	245
	Mediaster aequalis	Benthic	116
	Nearchaster (Nearchaster) aciculosus	Benthic	56
	Orthasterias koehleri	Benthic	102
	Pisaster brevispinus	Benthic	169
	Poraniopsis inflata	Benthic	91
	Pteraster tesselatus	Benthic	176
	Pycnopodia helianthoides	Benthic	434
	Sagenaster evermanni	Benthic	76
	Solaster dawsoni	Benthic	44
	Stylasterias forreri	Benthic	400
	Thrissacanthias penicillatus	Benthic	38
Echinoids (sea urchins; 2 spp.)	*Strongylocentrotus fragilis*	Benthic	1064
	Strongylocentrotus pallidus	Benthic	123
Holothuroids (sea cucumbers; 7 spp.)	*Apostichopus californicus*	Benthic	138
	Apostichopus leukothele	Benthic	244
	Capheira mollis	Benthic	41
	Pannychia moseleyi	Benthic	84
	Pseudostichopus mollis	Benthic	176
	Psolus squamatus	Benthic	38
	Synallactes challengeri	Benthic	43
Ophiuroids (brittle stars; 5 spp.)	*Asteronyx loveni*	Benthic	45
	Gorgonocephalus eucnemis	Benthic	198
	Ophiomusium lymani	Benthic	39
	Ophiura sarsii	Benthic	149

Phylum	Species	Habitat	Number of sites present
Ophiuroids (cont'd)	*Stegophiura ponderosa*	Benthic	70
Bivalves (2 spp.)	*Chlamys hastata*	Benthic	75
	Chlamys rubida	Benthic	145
Cephalopods (8 spp.)	*Chiroteuthis calyx*	Benthic	58
	Enteroctopus dofleini	Benthic	83
	Galiteuthis phyllura	Benthic	71
	Graneledone boreopacifica	Benthic	39
	Octopoteuthis deletron	Benthic	96
	Onykia robusta	Benthic	57
	Opisthoteuthis californiana	Benthic	97
	Rossia pacifica	Benthic	475
Gastropods (3 spp.)	*Armina californica*	Benthic	46
	Fusitriton oregonensis	Benthic	501
	Tochuina tetraquetra	Benthic	59
Total (174 species)			3615

Table D 2. List of species considered for use in cluster analysis, but excluded due to low frequency across study sites (i.e., these species belonged to an acceptable taxon with an acceptable habitat and were recorded in the Groundfish, Crab, or Tanner Crab research surveys, but were not represented in the final analysis). Threshold for inclusion was species found in less than 1% of sites (37 sites). Habitats were obtained from FishBase (2016), SeaLifeBase (2016), and other sources and were accurate from those sources as of December 2014.

Phylum	Species	Habitat	Number of sites present
Cirripedes (barnacles; 4 spp.)	*Balanus glandula*	Benthic	11
	Balanus nubilus	Benthic	16
	Chirona evermanni	Benthic	1
	Pollicipes polymerus	Benthic	5
Decapods (crabs, shrimp; 59 spp.)	*Acantholithodes hispidus*	Benthic	16
	Argis dentata	Benthic	1
	Argis lar	Benthic	26
	Argis ovifer	Benthic	5
	Betaeus setosus	Benthic	1
	Calocarides quinqueseriatus	Benthic	11
	Cancer productus	Benthic	32
	Chionoecetes bairdi	Benthic	16
	Crangon dalli	Benthic	27
	Crangon franciscorum franciscorum	Benthic	1
	Crangon nigricauda	Benthic	3
	Cryptolithodes typicus	Benthic	2
	Eualus barbatus	Benthic	20
	Glebocarcinus oregonensis	Benthic	10
	Hapalogaster mertensii	Benthic	1
	Heptacarpus flexus	Benthic	3
	Heptacarpus moseri	Benthic	4
	Heptacarpus tridens	Benthic	6
	Hyas lyratus	Benthic	4
	Lebbeus washingtonianus	Benthic	11
	Lithodes aequispinus	Benthic	35
	Lophopanopeus bellus	Benthic	1
	Metacarcinus gracilis	Benthic	7
	Metacrangon spinosissima	Benthic	2
	Metacrangon variabilis variabilis	Benthic	1
	Munidopsis quadrata	Benthic	8
	Neocrangon abyssorum	Benthic	17
	Oedignathus inermis	Benthic	1
	Oregonia bifurca	Benthic	23
	Pachycheles pubescens	Benthic	1
	Paguristes turgidus	Benthic	2

Phylum	Species	Habitat	Number of sites present
Decapods (cont'd)	*Pagurus aleuticus*	Benthic	3
	Pagurus beringanus	Benthic	1
	Pagurus brandti	Benthic	2
	Pagurus confragosus	Benthic	7
	Pagurus cornutus	Benthic	2
	Pagurus ochotensis	Benthic	1
	Pagurus stevensae	Benthic	7
	Pandalopsis ampla	Benthic	10
	Pandalopsis glabra	Benthic	7
	Pandalus goniurus	Benthic	11
	Pandalus hypsinotus	Benthic	21
	Pandalus stenolepis	Benthic	7
	Paralithodes brevipes	Benthic	27
	Paralithodes camtschaticus	Benthic	24
	Paralomis verrilli	Benthic	22
	Parapagurus benedicti	Benthic	11
	Parapasiphae sulcatifrons	Benthic	2
	Petrolisthes eriomerus	Benthic	2
	Phyllolithodes papillosus	Benthic	3
	Placetron wosnessenskii	Benthic	2
	Pugettia gracilis	Benthic	2
	Pugettia producta	Benthic	1
	Rhinolithodes wosnessenskii	Benthic	1
	Romaleon antennarium	Benthic	2
	Scyra acutifrons	Benthic	2
	Spirontocaris arcuata	Benthic	1
	Spirontocaris holmesi	Benthic	1
	Spirontocaris lamellicornis	Benthic	9
Fish (111 spp.)	*Acantholiparis opercularis*	Demersal	2
	Acipenser medirostris	Demersal	4
	Agonopsis vulsa	Demersal	28
	Alepocephalus tenebrosus	Demersal	32
	Allocareproctus jordani	Demersal	24
	Allosmerus elongatus	Demersal	20
	Amblyraja badia	Demersal	10
	Anoplagonus inermis	Demersal	9
	Aptocyclus ventricosus	Benthopelagic	2
	Aulorhynchus flavidus	Benthopelagic	3
	Bathyagonus alascanus	Demersal	13
	Bathyagonus infraspinatus	Demersal	1

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Bathymaster signatus*	Demersal	13
	Bathyraja maculata	Demersal	1
	Bathyraja minispinosa	Demersal	2
	Bothrocara molle	Demersal	22
	Bothrocara pusillum	Demersal	11
	Bryozoichthys marjorius	Demersal	29
	Careproctus colletti	Demersal	5
	Careproctus cypselurus	Demersal	5
	Careproctus furcellus	Demersal	17
	Careproctus ovigerus	Demersal	4
	Chesnonia verrucosa	Demersal	5
	Chirolophis decoratus	Demersal	2
	Citharichthys stigmaeus	Demersal	36
	Clidoderma asperrimum	Demersal	1
	Coryphaenoides leptolepis	Demersal	15
	Coryphaenoides liocephalus	Benthopelagic	2
	Coryphaenoides yaquinae	Demersal	4
	Cryptacanthodes aleutensis	Demersal	5
	Cryptacanthodes giganteus	Demersal	7
	Dasycottus setiger	Demersal	12
	Derepodichthys alepidotus	Demersal	5
	Elassodiscus caudatus	Demersal	29
	Eptatretus stoutii	Demersal	26
	Erilepis zonifer	Demersal	7
	Eumicrotremus orbis	Demersal	13
	Gobiesox maeandricus	Demersal	2
	Hemilepidotus spinosus	Demersal	21
	Hexagrammos stelleri	Demersal	3
	Hexanchus griseus	Demersal	14
	Hypomesus pretiosus	Benthopelagic	1
	Hypsagonus quadricornis	Demersal	1
	Icelinus borealis	Demersal	15
	Icelinus burchami	Demersal	15
	Icelus spiniger	Demersal	6
	Jordania zonope	Demersal	1
	Lampetra ayresii	Demersal	3
	Leptoclinus maculatus	Demersal	19
	Leptocottus armatus	Demersal	29
	Limanda aspera	Demersal	3
	Liparis fucensis	Demersal	2

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Liparis gibbus*	Demersal	1
	Lipariscus nanus	Benthopelagic	1
	Lycenchelys camchatica	Demersal	2
	Lycenchelys jordani	Demersal	9
	Lycenchelys micropora	Demersal	11
	Lycodapus dermatinus	Demersal	4
	Malacocottus aleuticus	Benthopelagic	22
	Myoxocephalus polyacanthocephalus	Demersal	22
	Nautichthys oculofasciatus	Demersal	5
	Odontopyxis trispinosa	Demersal	10
	Oncorhynchus mykiss	Benthopelagic	3
	Pachycara lepinium	Demersal	3
	Pallasina barbata	Demersal	1
	Paraliparis cephalus	Demersal	10
	Paraliparis deani	Demersal	3
	Paraliparis melanobranchus	Demersal	1
	Paraliparis paucidens	Demersal	8
	Paraliparis pectoralis	Demersal	7
	Paraliparis rosaceus	Demersal	33
	Paraliparis ulochir	Demersal	1
	Paricelinus hopliticus	Demersal	9
	Pholis laeta	Demersal	1
	Pholis ornata	Demersal	1
	Pholis schultzi	Demersal	2
	Pleuronichthys coenosus	Demersal	6
	Porichthys notatus	Demersal	4
	Psychrolutes phrictus	Demersal	22
	Psychrolutes sigalutes	Demersal	10
	Ptilichthys goodei	Demersal	1
	Radulinus boleoides	Demersal	1
	Radulinus taylori	Demersal	3
	Rhacochilus vacca	Demersal	6
	Rhamphocottus richardsonii	Demersal	6
	Rhinoliparis attenuatus	Demersal	12
	Rhinoliparis barbulifer	Demersal	2
	Scorpaenichthys marmoratus	Demersal	37
	Sebastes auriculatus	Demersal	4
	Sebastes chlorostictus	Demersal	3
	Sebastes emphaeus	Reef-associated	36
	Sebastes goodei	Demersal	5

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Sebastes melanops*	Reef-associated	24
	Sebastes melanostomus	Demersal	13
	Sebastes mystinus	Reef-associated	2
	Sebastes polyspinis	Demersal	1
	Sebastes rufus	Demersal	3
	Sebastes saxicola	Demersal	31
	Sebastes semicinctus	Demersal	1
	Sebastes variabilis	Demersal	33
	Spectrunculus grandis	Demersal	2
	Spirinchus starksi	Benthopelagic	1
	Spirinchus thaleichthys	Benthopelagic	1
	Symphurus atricaudus	Demersal	1
	Talismania bifurcata	Demersal	26
	Trichodon trichodon	Demersal	26
	Triglops pingelii	Demersal	32
	Xeneretmus triacanthus	Demersal	23
	Xiphister mucosus	Demersal	1
	Zaprora silenus	Demersal	25
	Zesticelus profundorum	Demersal	3
Tunicates (2 spp.)	*Chelyosoma productum*	Benthic	9
	Halocynthia hilgendorfi	Benthic	7
Anthozoans (7 spp.)	*Anthoptilum grandiflorum*	Benthic	2
	Bathypathes patula	Benthic	3
	Cymbactis faeculenta	Benthic	1
	Primnoa pacifica	Benthic	1
	Stylatula elongata	Benthic	6
	Swiftia pacifica	Benthic	1
	Urticina columbiana	Benthic	1
Asteroids (sea stars; 53 spp.)	*Ampheraster marianus*	Benthic	9
	Asthenactis fisheri	Benthic	2
	Astropecten armatus	Benthic	1
	Benthopecten claviger	Benthic	1
	Ceramaster arcticus	Benthic	3
	Ceramaster clarki	Benthic	2
	Ceramaster japonicus	Benthic	6
	Diplopteraster multiples	Benthic	21
	Dipsacaster anoplus	Benthic	3
	Dipsacaster borealis	Benthic	12
	Evasterias troschelii	Benthic	6
	Freyella microplax	Benthic	1

Phylum	Species	Habitat	Number of sites present
Asteroids (cont'd)	*Gephyreaster swifti*	Benthic	12
	Henricia aspera	Benthic	32
	Henricia asthenactis	Benthic	14
	Henricia leviuscula	Benthic	12
	Henricia leviuscula annectens	Benthic	4
	Henricia leviuscula spiculifera	Benthic	4
	Henricia longispina	Benthic	22
	Henricia polyacantha	Benthic	2
	Henricia sanguinolenta	Benthic	30
	Hippasteria lepidonotus	Benthic	3
	Hymenaster pellucidus	Benthic	7
	Hymenaster quadrispinosus	Benthic	15
	Hymenodiscus pannychia	Benthic	9
	Hymenodiscus pusilla	Benthic	7
	Leptasterias hexactis	Benthic	1
	Leptychaster anomalus	Benthic	24
	Leptychaster arcticus	Benthic	10
	Leptychaster pacificus	Benthic	7
	Mediaster tenellus	Benthic	18
	Myxoderma sacculatum	Benthic	11
	Nearchaster (Nearchaster) variabilis	Benthic	20
	Patiria miniata	Benthic	1
	Pectinaster agassizi evoplus	Benthic	13
	Pedicellaster magister	Benthic	11
	Pisaster giganteus	Benthic	3
	Pisaster ochraceus	Benthic	9
	Pseudarchaster alascensis	Benthic	31
	Pseudarchaster dissonus	Benthic	23
	Pseudarchaster parelii	Benthic	6
	Psilaster pectinatus	Benthic	21
	Pteraster jordani	Benthic	22
	Pteraster marsippus	Benthic	4
	Pteraster militaris	Benthic	10
	Pteraster trigonodon	Benthic	1
	Rathbunaster californicus	Benthic	25
	Solaster endeca	Benthic	3
	Solaster hypothryssus	Benthic	11
	Solaster paxillatus	Benthic	33
	Solaster stimpsoni	Benthic	17
	Tarsaster alaskanus	Benthic	24

Phylum	Species	Habitat	Number of sites present
Asteroids (cont'd)	*Zoroaster ophiurus*	Benthic	1
Crinoids (sea lilies; 4 spp.)	*Florometra asperrima*	Benthic	19
	Florometra serratissima	Benthic	16
	Psathyrometra fragilis	Benthic	9
	Ptilocrinus (Ptilocrinus) pinnatus	Benthic	4
Echinoids (sea urchins; 3 spp.)	*Mesocentrotus franciscanus*	Benthic	14
	Strongylocentrotus droebachiensis	Benthic	16
	Strongylocentrotus purpuratus	Benthic	10
Holothuroids (sea cucumbers; 9 spp.)	*Cucumaria frondosa japonica*	Benthic	1
	Cucumaria miniata	Benthic	3
	Cucumaria piperata	Benthic	2
	Eupentacta quinquesemita	Benthic	16
	Molpadia intermedia	Benthic	37
	Paracaudina chilensis	Benthic	4
	Pentamera lissoplaca	Benthic	3
	Pentamera pseudocalcigera	Benthic	1
	Psolus chitonoides	Benthic	8
Ophiuroids (brittle stars, 18 spp.)	*Amphiophiura superba*	Benthic	33
	Amphioplus (Amphioplus) strongyloplax	Benthic	6
	Amphiura (Amphiura) diomedeae	Benthic	11
	Asteroschema sublaeve	Benthic	10
	Ophiacantha bathybia	Benthic	1
	Ophiacantha rhachophora	Benthic	2
	Ophiomusium glabrum	Benthic	18
	Ophiopholis aculeata	Benthic	6
	Ophiopholis bakeri	Benthic	2
	Ophiopholis japonica	Benthic	4
	Ophiopholis longispina	Benthic	15
	Ophiophthalmus cataleimmoidus	Benthic	10
	Ophiophthalmus normani	Benthic	36
	Ophiopleura borealis	Benthic	3
	Ophioscolex corynetes	Benthic	4
	Ophiosphalma jolliense	Benthic	25
	Ophiura flagellata	Benthic	1
	Ophiura luetkenii	Benthic	1
Bivalves (25 spp.)	*Acharax johnsoni*	Benthic	7
	Cardiomya planetica	Benthic	1
	Compsomyax subdiaphana	Benthic	1
	Crassadoma gigantea	Benthic	5

Phylum	Species	Habitat	Number of sites present
Bivalves (cont'd)	*Dallicordia alaskana*	Benthic	1
	Delectopecten vancouverensis	Benthic	14
	Dermatomya tenuiconcha	Benthic	8
	Entodesma navicula	Benthic	1
	Halicardia perplicata	Benthic	1
	Leukoma staminea	Benthic	1
	Macoma nasuta	Benthic	3
	Mactromeris polynyma	Benthic	1
	Malletia faba	Benthic	3
	Malletia pacifica	Benthic	1
	Megayoldia montereyensis	Benthic	2
	Mytilus edulis	Benthic	1
	Nucula carlottensis	Benthic	1
	Nuculana conceptionis	Benthic	1
	Nuculana leonina	Benthic	1
	Panomya ampla	Benthic	1
	Patinopecten caurinus	Benthic	3
	Phreagena kilmeri	Benthic	2
	Pododesmus macrochisma	Benthic	4
	Policordia jeffreysi	Benthic	1
	Vesicomya stearnsii	Benthic	2
Cephalopods (9 spp.)	*Architeuthis dux*	Benthic	8
	Benthoctopus leioderma	Benthic	23
	Berryteuthis anonychus	Benthic	4
	Gonatopsis borealis	Benthic	7
	Gonatus berryi	Benthic	2
	Gonatus onyx	Benthic	2
	Gonatus pyros	Benthic	1
	Octopus rubescens	Benthic	16
	Ommastrephes bartramii	Benthopelagic	8
Gastropods (13 spp.)	*Arctomelon stearnsii*	Benthic	1
	Barleeia haliotiphila	Benthic	1
	Barleeia subtenuis	Benthic	1
	Bathybembix bairdii	Benthic	2
	Calliostoma platinum	Benthic	6
	Cidarina cidaris	Benthic	17
	Doris odhneri	Benthic	14
	Euspira pallida	Benthic	1
	Margarites helicinus	Benthic	1
	Margarites vorticiferus	Benthic	1

Phylum	Species	Habitat	Number of sites present
Gastropods (cont'd)	*Neptunea amianta*	Benthic	3
	Neverita lewisii	Benthic	12
	Triopha catalinae	Benthic	2
Polyplacophorans (chitons; 1 spp.)	*Cryptochiton stelleri*	Benthic	1
Total (318 species)			92

Table D 3. List of species excluded from analysis because of a pelagic, freshwater, or unknown habitat type (i.e., these species belonged to an acceptable taxon and were recorded in the Groundfish, Crab, or Tanner Crab research surveys, but were not represented in the final analysis). Habitats were obtained from FishBase (2016),SeaLifeBase (2016), and other sources and were accurate from those sources as of December 2014. The category "unknown" was assigned if no accurate information on a species' habitat could be found, or if a species has life history stages with different habitats that were not obvious in the original data.

Phylum	Species	Habitat	Number of sites present
Decapods (crabs, shrimp; 8 spp.)	*Acanthephyra curtirostris*	Pelagic	13
	Bentheogennema borealis	Pelagic	16
	Bentheogennema burkenroadi	Pelagic	30
	Calastacus stilirostris	Unknown	4
	Eusergestes similis	Pelagic	30
	Hymenodora frontalis	Pelagic	39
	Lophaxius rathbunae	Unknown	12
	Systellaspis braueri braueri	Pelagic	45
Fish (80 spp.)	*Acanthonus armatus*	Pelagic	1
	Alosa sapidissima	Pelagic	62
	Anoplogaster cornuta	Pelagic	4
	Anotopterus nikparini	Pelagic	5
	Antimora microlepis	Pelagic	340
	Aphanopus carbo	Pelagic	1
	Arctozenus risso	Pelagic	15
	Argyropelecus sladeni	Pelagic	2
	Aristostomias scintillans	Pelagic	37
	Avocettina infans	Pelagic	99
	Bathophilus flemingi	Pelagic	24
	Bathylagus pacificus	Pelagic	95
	Bathylychnops exilis	Pelagic	3
	Benthalbella dentata	Pelagic	56
	Benthalbella linguidens	Pelagic	5
	Brama japonica	Pelagic	2
	Chauliodus macouni	Pelagic	295
	Clupea pallasii pallasii	Pelagic	558
	Cololabis saira	Pelagic	6
	Coryphaenoides armatus	Pelagic	14
	Cottus cognatus	Freshwater	1
	Cyclothone atraria	Pelagic	6
	Diaphus theta	Pelagic	213
	Engraulis mordax	Pelagic	4
	Galeorhinus galeus	Pelagic	55
	Halargyreus johnsonii	Pelagic	11

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Holtbyrnia latifrons*	Pelagic	1
	Icichthys lockingtoni	Pelagic	7
	Icosteus aenigmaticus	Pelagic	15
	Kali indica	Pelagic	1
	Lamna ditropis	Pelagic	9
	Lampadena urophaos urophaos	Pelagic	9
	Lampanyctus jordani	Pelagic	20
	Lestidiops ringens	Pelagic	8
	Leuroglossus schmidti	Pelagic	11
	Lycodapus endemoscotus	Pelagic	8
	Lycodapus fierasfer	Pelagic	70
	Lycodapus mandibularis	Pelagic	42
	Lycodapus pachysoma	Pelagic	21
	Macropinna microstoma	Pelagic	17
	Magnisudis atlantica	Pelagic	5
	Melamphaes lugubris	Pelagic	47
	Melanostigma pammelas	Pelagic	9
	Merluccius productus	Pelagic	1106
	Nannobrachium regale	Pelagic	159
	Nannobrachium ritteri	Pelagic	88
	Nansenia candida	Pelagic	24
	Nectoliparis pelagicus	Pelagic	7
	Nemichthys larseni	Pelagic	4
	Nemichthys scolopaceus	Pelagic	16
	Notacanthus chemnitzii	Pelagic	10
	Oncorhynchus gorbuscha	Pelagic	61
	Oncorhynchus kisutch	Pelagic	56
	Oncorhynchus nerka	Pelagic	17
	Oneirodes bulbosus	Pelagic	5
	Oneirodes thompsoni	Pelagic	4
	Polyacanthonotus challengeri	Pelagic	3
	Poromitra crassiceps	Pelagic	124
	Prionace glauca	Pelagic	131
	Protomyctophum thompsoni	Pelagic	20
	Pseudobathylagus milleri	Pelagic	129
	Sagamichthys abei	Pelagic	46
	Sardinops sagax	Pelagic	50
	Scomber japonicus	Pelagic	3
	Scopelengys tristis	Pelagic	9
	Scopeloberyx robustus	Pelagic	12

Phylum	Species	Habitat	Number of sites present
Fish (cont'd)	*Scopelosaurus harryi*	Pelagic	32
	Sebastes entomelas	Pelagic	289
	Serrivomer jesperseni	Pelagic	4
	Squalus suckleyi	Pelagic	2324
	Stenobrachius leucopsarus	Pelagic	316
	Stenobrachius nannochir	Pelagic	132
	Sternoptyx pseudobscura	Pelagic	1
	Symbolophorus californiensis	Pelagic	53
	Tactostoma macropus	Pelagic	128
	Tarletonbeania crenularis	Pelagic	108
	Thalassenchelys coheni	Pelagic	1
	Thaleichthys pacificus	Pelagic	600
	Trachipterus altivelis	Pelagic	1
	Trachurus symmetricus	Pelagic	36
Tunicates (3 spp.)	*Cyclosalpa affinis*	Pelagic	202
	Pegea confoederata	Pelagic	42
	Salpa maxima	Pelagic	80
Hydrozoans (2 spp.)	*Dimophyes arctica*	Unknown	1
	Sertularella tanneri	Unknown	1
Scyphozoans (6 spp.)	*Aurelia aurita*	Unknown	21
	Chrysaora melanaster	Unknown	4
	Chrysaora quinquecirrha	Pelagic	2
	Cyanea capillata	Unknown	311
	Periphylla periphylla	Unknown	60
	Phacellophora camtschatica	Unknown	9
Cephalopods (11 spp.)	*Abraliopsis felis*	Pelagic	8
	Berryteuthis magister	Pelagic	471
	Doryteuthis opalescens	Pelagic	207
	Dosidicus gigas	Pelagic	34
	Histioteuthis heteropsis	Pelagic	1
	Histioteuthis hoylei	Pelagic	2
	Japetella diaphana	Pelagic	48
	Muusoctopus robustus	Unknown	1
	Onychoteuthis borealijaponica	Pelagic	3
	Taonius borealis	Unknown	4
	Vampyroteuthis infernalis	Pelagic	11
Total (110 species)			3221

APPENDIX E. TRANSITION ZONE DIVERSITY ANALYSIS

METHODS

The following procedure was carried out for both of the biophysical unit analyses (4a and 4b). All sites with a random forest model probability of assignment ≥ 0.70 were considered "biophysical unit sites", whereas those with probabilities < 0.70 were considered "transition sites". To compare the species richness between transition and biophysical unit sites, it was necessary to use a rarefaction method to account for the larger number of sites classified as biophysical unit sites compared to transition sites. Using function *specaccum* in the R package 'vegan', species accumulation curves were created for transition and biophysical unit sites for each full dataset, and considering each biophysical unit separately. *Speccacum* uses random permutations of the site by species matrix to calculate the mean number of species observed at a given number of sites. For each analysis, the number of observed species in the transition sites was compared to the expected number of species (ES_n) at the n-th biophysical unit site, where n is the number of transition sites.

LEVEL 4A BIOPHYSICAL UNITS

Results shown in Table E 1 and Figure E 1. Overall the large scale biophysical unit (level 4a) sites have a higher expected number of species than the transition sites (ES_{70} = 151 ± 6 vs 140 ± 0). Considered individually, both the Banks and Slope biophysical units have fewer species than observed in adjoining transition sites (Banks: ES_{39} = 94 ± 0 vs 78 ± 4; Slope: ES_{10} = 74 ± 0 vs 45 ± 11), while the opposite is true for the shelf biophysical unit (ES_{10} = 90 ± 8 in non-transition sites, vs 70 ± 0 for transition sites).

LEVEL 4B BIOPHYISCAL UNITS

Results shown in Table E 1Table E 1 and Figure E 2. There were more transition sites identified in the level 4b biophysical unit analysis (191) than in the level 4a biophysical unit analysis (70). As for level 4a, the overall comparison shows more species in the biophysical unit sites than the transition sites (ES_{191} = 171 ± 2 vs 155 ± 0). However, when examined individually, the Other Banks, Dogfish Banks, Shelf, and Slope biophysical units had fewer species present compared to their corresponding transition sites in the. The Troughs biophysical unit sites had roughly the same number of species as the adjoining transition sites (ES_{69} = 112 ± 6 vs 114 ± 0).

*Table E 1. ES_n (expected number of species) for biophysical unit sites (random forest probability of assignment ≥ 0.70), and transition sites (probability of assignment < 0.70), where n is the number of sites in either biophysical unit or transition sections. * indicates the higher ES_n between transition or biophysical unit.*

Analysis	Subset	n biophysical unit sites	n transition sites	ES_n	ES_n at biophysical unit sites	ES_n at transition sites
Level 4a biophysical units	All	3541	70	ES_{70}	151 ± 6 *	140 ± 0
	Banks	260	39	ES_{39}	78 ± 4	94 ± 0 *
	Shelf	2704	21	ES_{21}	90 ± 8 *	70 ± 0
	Slope	578	10	ES_{10}	45 ± 11	74 ± 0 *
Level 4b biophysical units	All	3241	191	ES_{191}	171 ± 2 *	155 ± 0
	Other Banks	75	45	ES_{45}	82 ± 3	100 ± 0 *
	Dogfish Bank	157	23	ES_{23}	61 ± 4	73 ± 0 *
	Shelf	1766	43	ES_{43}	100 ± 6	118 ± 0 *
	Slope	575	12	ES_{12}	54 ± 11	77 ± 0 *
	Troughs	848	69	ES_{69}	112 ± 6	114 ± 0

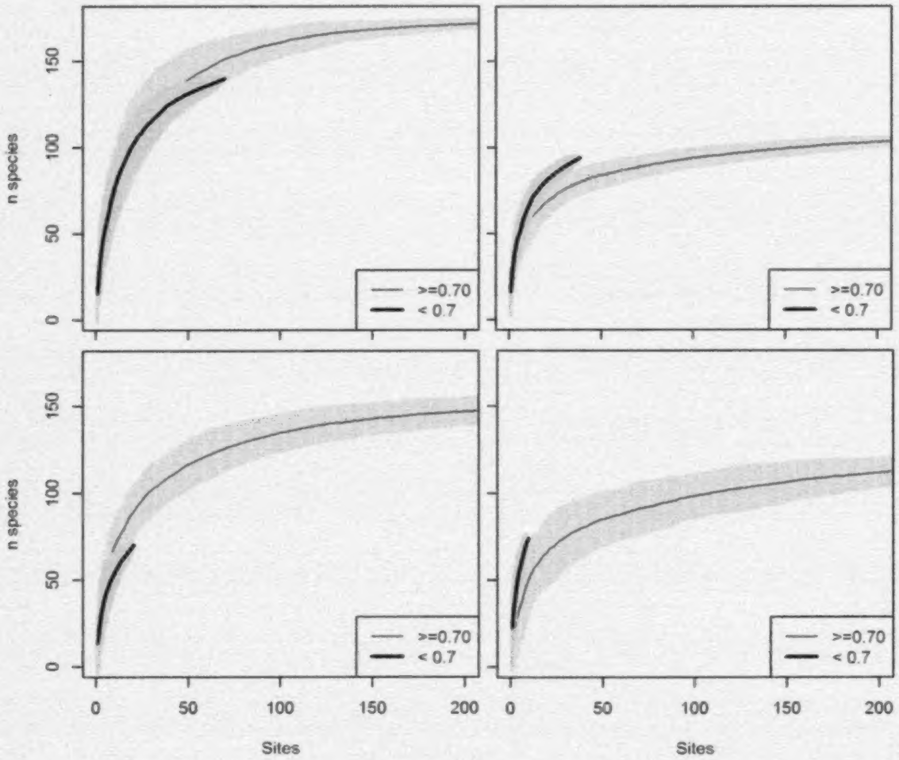


Figure E 1. Species accumulation curves for the level 4a biophysical unit analysis. Blue lines show accumulated number of species in biophysical unit sites (random forest probability of assignment ≥ 0.70), while black lines show accumulated number of species in transition sites (probability of assignment < 0.70). X-axes (number of sites) shows only first 200 sites (see Table Y for actual number of sites) to improve visualization of the smaller transition site lines. A) All level 4a biophysical unit data, b) Banks biophysical unit, c) Shelf biophysical unit, d) Slope biophysical unit.

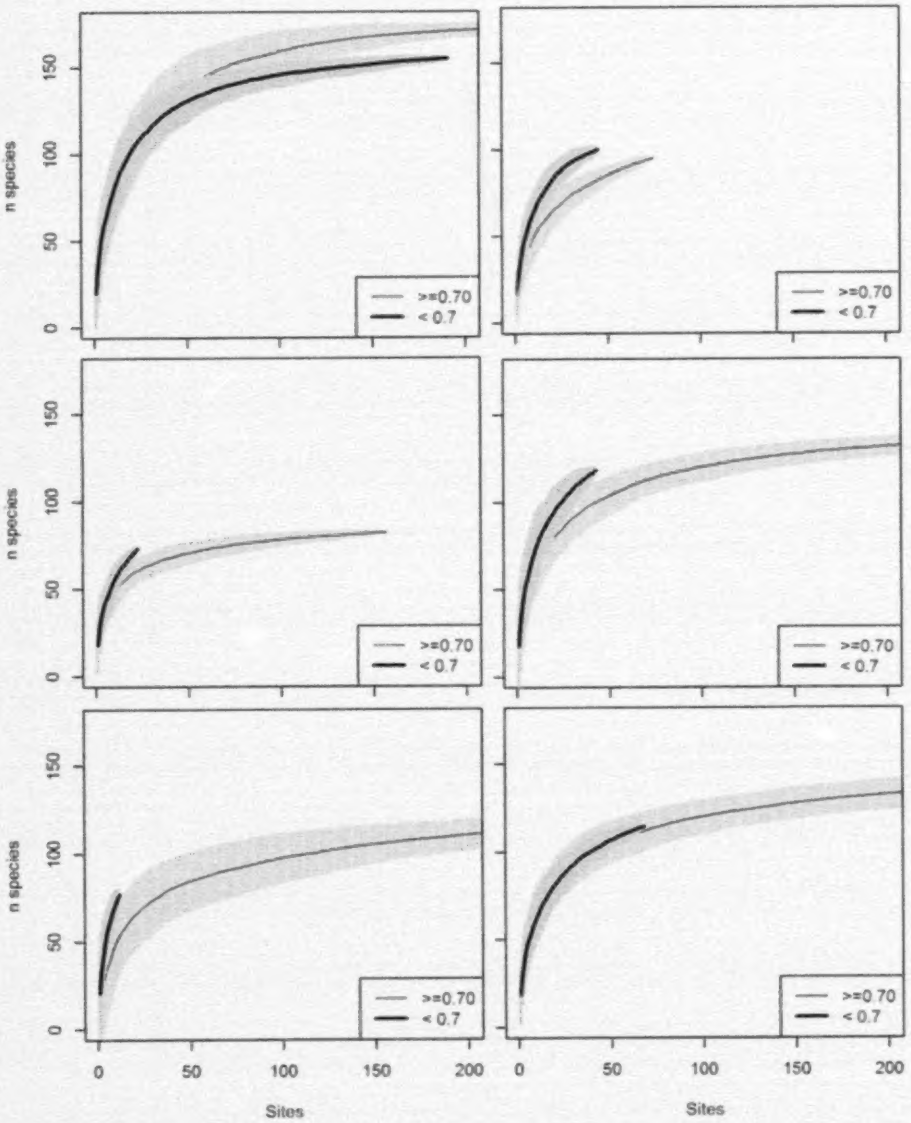


Figure E 2. Species accumulation curves for the level 4b biophysical unit analysis. Blue lines show accumulated number of species in biophysical unit sites (random forest probability of assignment ≥ 0.70), while black lines show accumulated number of species in transition sites (probability of assignment < 0.70). X-axes (number of sites) shows only first 200 sites (see Table Y for actual number of sites) to improve visualization of the smaller transition site lines. a) All level 4b biophysical unit data, b) Other banks biophysical unit, c) Dogfish Bank biophysical unit, d) Shelf biophysical unit, e) Slope biophysical unit, F) Troughs biophysical unit.

APPENDIX F. GEOMORPHIC UNITS

GEOMORPHIC UNIT ATTRIBUTES

Geomorphic Unit – Geomorphic unit class

Analysis Area – the analysis area associated with the geomorphic unit. Separate analyses were performed for the continental slope, shelf, and fjords.

AssocGazette – the gazetted names of undersea features that overlap the geomorphic unit. There may be more than one gazetted feature associated with each geomorphic unit polygon. For canyons and troughs, the associated gazetted feature was added to all of the canyon and trough bottom geomorphic units that fell within the boundaries of the canyon or trough polygon identified by Manson (2009) that shared the same gazetted name.

Geozone_Greene– geomorphic unit class described using the classification developed by Greene et al. (2008).

Table F 1. Classification scheme for geological features and zones after Manson (2009).

Term	Description
Seamounts, hills, ridges on the continental rise	Areas with slope >3°; BPI >100
Canyons and valleys on the continental slope	Depressions with BPI <-50; steep-sided, with slope >4 over a significant portion of the canyon sides; narrow, with ~50% of the width of the feature being the steep sides; deep, with the height of the sides being ~20% of the width. Features with bottoms of continuous gradient were valleys, others were designated canyons.
Troughs, valleys and basis on the continental shelf	Depressions with BPI <0 and low gradient bottoms with no major breaks and slope <2.
Troughs, valleys and basins on the continental rise	Depressions with BPI <0 and low gradient bottoms with no major breaks and slope <2 on the continental rise.

Table F 2. Classification scheme for geological features and zones after Ferdaña (2006).

Term	BPI	Slope	Depth
Inner shelf ridge	> mean + 1 SD	-	< 40 m
Mid shelf ridge	> mean + 1 SD	-	40.1–200 m
Mesobenthal ridge	> mean + 1 SD	-	200.1–700 m
Bathybenthal ridge	> mean + 1 SD	-	700.1–5000 m
Inner shelf slope	≥ -1 SD, ≤ 1 SD	> 5°	< 40 m
Mid shelf slope	≥ -1 SD, ≤ 1 SD	> 5°	40.1–200 m
Mesobenthal slope	≥ -1 SD, ≤ 1 SD	> 5°	200.1–700 m
Bathybenthal slope	≥ -1 SD, ≤ 1 SD	> 5°	700.1–5000 m
Inner shelf flats	≥ -0.5 SD, ≤ 0.5 SD	≤ 5°	< 40 m
Mid shelf flats	≥ -0.5 SD, ≤ 0.5 SD	≤ 5°	40.1–200 m
Mesobenthal flats	≥ -0.5 SD, ≤ 0.5 SD	≤ 5°	200.1–700 m
Bathybenthal flats	≥ -0.5 SD, ≤ 0.5 SD	≤ 5°	700.1–5000 m
Inner shelf canyon	< -1 SD	-	< 40 m
Mid shelf canyon	< -1 SD	-	40.1–200 m
Mesobenthal canyon	< -1 SD	-	200.1–700 m
Bathybenthal canyon	< -1 SD	-	700.1–3500 m

LITERATURE CITED - APPENDICES

Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26: 32-46

Anderson, M.J. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62: 245-253

Anderson, M.J., and Walsh, D.C.I. 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecol. Monog. 83(4): 557-574

Choromanski, E.M., Workman, G.D., and Fargo, J. 2005. Hecate Strait Multi-species bottom trawl survey, CCGS W.E. Ricker, May 19 to June 7, 2003. Can. Data Rep. Fish. Aquat. Sci. 1169: 85 p

DFO Crab by Trap IFMP. 2015. Pacific Region integrated fisheries management plan: crab by trap, January 1 to December 31, 2015.

DFO Groundfish IFMP. 2011. Pacific Region integrated fisheries management plan: groundfish, February 21, 2011 to February 20, 2013.

Dunham, J.S., Phillips, A., Morrison, J., and Jorgensen, G. 2011. A manual for Dungeness crab surveys in British Columbia. Can. Tech. Rep. Fish. Aquat. Sci. 2964: viii + 68 p.

Ferdaña, Z. 2006. Case study: Pacific Northwest Coast (PNWC) ecoregion offshore classification methodology. *In* Ecoregional Assessment and Biodiversity Vision Toolbox *Edited by* J. Higgins and R. Esselman. The Nature Conservancy, Arlington, VA.

FishBase. 2016. World Wide Web electronic publication. Froese, R. and Pauly, D. (eds). (Accessed 15 March 2016)

Fleischer, G.W., Cooke, K.D., Ressler, P.H., Thomas, R.E., de Blois, S.K., Hufnagle, L.C., Kronlund, A.R., Holmes, J.A., and Wilson, C.D. 2005. The 2003 Integrated Acoustic and Trawl Survey of Pacific Hake, Merluccius productus, in U.S. and Canadian Waters off the Pacific Coast. U.S. Dept. Commer. NOAA Tech. Memo. NMFS-NWFSC-65: 45 p

Flemming, R.G., Yamanaka, K.L., Cooke, K., and Dykstra, C. 2012. Summary of Non-Halibut Catch from the Standardized Stock Assessment Survey Conducted by the International Pacific Halibut Commission in British Columbia from June 3 to August 27, 2010. Can. Tech. Rep. Fish. Aquat. Sci. 2989

Gillespie, G.E., Fong, K.H., Phillips, A.C., Meyer, G.R., and Boutillier, J.A. 2004. Development of a New Fishery for Tanner Crabs (Chionoecetes tanneri Rathbun, 1893) off British Columbia: 2003 Status Report. DFO Can. Sci. Advis. Sec. Res. Doc. 2004/132. viii + 87 p.

Greene, H.G., O'Connell, V., Brylinsky, C.K., and Reynolds, J.R. 2008. Marine Benthic Habitat Classification: What's Best for Alaska? *In* Marine Habitat Mapping Technology for Alaska. *Edited by* J.R. Reynolds and H.G. Greene. Alaska Sea Grant College Program, University of Alaska Fairbanks, Fairbanks, Alaska. pp. 169-184.

King, J.R., McFarlane, G.A., Jones, S.R.M., Gilmore, S.R., and Abbott, C.L. 2012. Stock delineation of migratory and resident Pacific hake in Canadian waters. Fish. Res. 114: 19-30

Krishka, B.A., Starr, P.J., and Choromanski, E.M. 2005. Longspine Thornyhead Random Stratified Trawl Survey off the West Coast of Vancouver Island, September 4-21, 2003. Can. Tech. Rep. Fish. Aquat. Sci. 2577

Lochead, J.K., and Yamanaka, K.L. 2007. Summary Report for the inshore rockfish (*Sebastes* spp.) longline survey conducted in statistical areas 14 to 20, 28 and 29, from August 11 to September 6, 2005. Can. Tech. Rep. Fish. Aquat. Sci. 2690: 53 p

Manson, M.M. 2009. Small Scale Delineation of Northeast Pacific Ocean Undersea Features Using Benthic Position Index. Can. Manuscr. Rep. Fish. Aquat. Sci.2864: iv + 16 p.

McCune, B., and Grace, J. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon.

Oksanen, J. 2014. Vegan: ecological diversity.(Accessed 15 March 2016)

Olsen, N., Rutherford, K.L., Stanley, R.D., and Wyeth, M. 2009a. West Coast Vancouver Island Groundfish Bottom Trawl Survey, May 26th to June 22nd, 2008. Can. Man. Rep. Fish. Aquat. Sci. 2092: 50 p

Olsen, N., Rutherford, K.L., Stanley, R.D., and Wyeth, M.R. 2009b. Hecate Strait Groundfish Bottom Trawl Survey, May 26th to June 21st, 2009. Can. Man. Rep. Fish. Aquat. Sci. 2901: 49 p

SeaLifeBase. 2016. World Wide Web electronic publication. Palomares, M.L.D. and Pauly, D. (eds). (Accessed 15 March 2016)

Sinclair, A., and Workman, G. 2002. A review of Pacific cod (*Gadus macrocephalus*) monitoring surveys in Hecate Strait, March-July 2002. DFO Can. Sci. Advis. Sec. Res. Doc. 2002/130. 59 p.

Stanley, R.D., Olson, N., Workman, G.D., Cleary, J., and de la Mare, W. 2007. A review of the Queen Charlotte Sound Groundfish Bottom Trawl Survey (2003-2005). Can. Tech. Rep. Fish. Aquat. Sci. 2709

White, E., Soderlund, E., Dykstra, C.L., Geernaert, T.O., and Ranta, A.M. 2010. 2010 Standardized stock assessment survey. (Accessed 15 March 2016)

Workman, G.D., Olsen, N., Fargo, J., and Stanley, R.D. 2008a. West Coast Vancouver Island Groundfish Bottom Trawl Survey, R/V WE RICKER, May 25th to June 10th, 2004. Can. Man. Rep. Fish. Aquat. Sci. 2826: 64 p

Workman, G.D., Stanley, R.D., Olsen, N., and Rutherford, K.L. 2008c. West Coast Queen Charlotte Islands Groundfish Bottom Trawl Survey, September 11th to October 17th, 2007. Can. Man. Rep. Fish. Aquat. Sci. 2823: 45 p

Wyeth, M.R., Kronlund, A.R., and Elfert, M. 2007. Summary of the 2005 British Columbia sablefish (*Anoplopoma fimbria*) research and assessment survey. Can. Tech. Rep. Fish. Aquat. Sci. 2694

Yamanaka, K.L., Lacko, L.C., Lochead, J.K., Martin, J., Haigh, R., Grandin, C., and West, K. 2004a. Stock Assessment Framework for Inshore Rockfish. DFO Can. Sci. Advis. Sec. Res. Doc. 2004/068. iii + 63 p.

Yamanaka, K.L., Lochead, J., and Dykstra, C. 2004b. Summary of Non-Halibut Catch from the Standardized Stock Assessment Survey Conducted by the International Pacific Halibut Commission in British Columbia from May 27 to August 11, 2003. Can. Tech. Rep. Fish. Aquat. Sci. 2535

Zacharias, M.A., Howes, D.E., Harper, J.R., and Wainwright, P. 1998. The British Columbia marine ecosystem classification: Rationale, development, and verification. Coast. Manage. 26(2): 105-124